

-1199-

Query: 192 VRAHPIDELLKLYVRINAEAEETDPSVDEEAREWFRKLEANDPEATELWQWFRDESLLLEFN 251
 V+AHPIDELLKLYVRINAEAEETDP+VDEEAREWFRKLE D EATELWQWFRDESLLLEFN
 Sbjct: 181 VQAHPIDELLKLYVRINAEAEETDPTVDEEAREWFRKLEDGDKEATELWQWFRDESLLLEFN 240

5 Query: 252 RLYDQMNVTFDSDYNGEAFYNDKMDDEVLELLESKNLLVESKGAQVVNLEKYGIEHPALIKK 311
 RLYDQ++VTFDSYNGEAFYNDKMDDEV+LLE+KNLLVESKGAQVVNLEKYGIEHPALIKK
 Sbjct: 241 RLYDQLHVTFDSDYNGEAFYNDKMDDEVLDLLEAKNLLVESKGAQVVNLEKYGIEHPALIKK 300

10 Query: 312 SDGATLYITRDLAALYRKRTYDFAKSIYVVGNEQSAHFKQLKAVLKEMDYDWSDDMTHV 371
 SDGATLYITRDLAALYRKRTYDFAKS+YVVGNEQ+AHFKQLKAVLKEM YDWSDDMTHV
 Sbjct: 301 SDGATLYITRDLAALYRKRTYDFAKSVYVVGNEQAAHFKQLKAVLKEMGYDWSDDMTHV 360

15 Query: 372 PFGLVTKGGAKLSTRKGNVILLEPTVAEAINRAASQIEAKNPNLADKDKVAQAVGVGAIK 431
 FGLVTKGGAKLSTRKGNVILLEPTVAEAINRAASQIEAKNPNLADK+ VA AVGVGAIK
 Sbjct: 361 AFGLVTKGGAKLSTRKGNVILLEPTVAEAINRAASQIEAKNPNLADKEAVAHAVGVGAIK 420

20 Query: 432 FYDLKTDRTNGYDFDLEAMVSFEGETGPPYQYAHARIQSILRKANFSPSNSDNYSLNDVE 491
 FYDLKTDNR NGYDFDLEAMVSFEGETGPPYQYAHARIQSILRKA+F+PS + YSL D E
 Sbjct: 421 FYDLKTDNRNGYDFDLEAMVSFEGETGPPYQYAHARIQSILRKADFTPSATTTYSLADAE 480

25 Query: 492 SWEIKLIQDFPRIIVRAADNFEPSSI+AKFAINLAQCFNKYYAHTRILDEDAEISSRLAL 551
 SWEIKLIQDFPRII R +DNFEP+AKFAINLAQ FNKYYAHTRILD+++E +RLAL
 Sbjct: 481 SWEIKLIQDFPRIIKRTSDNFEP+AKFAINLAQSFNKYYAHTRILDDNSERDNRAL 540

Query: 552 CYATATVLKESLRLLGVDAPNEM 574
 CYATATVLKE+LRLLGVDAPNEM
 Sbjct: 541 CYATATVLKEALRLLGVDAPNEM 563

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1076

A DNA sequence (GBSx1150) was identified in *Sagalactiae* <SEQ ID 3315> which encodes the amino acid sequence <SEQ ID 3316>. This protein is predicted to be arginine hydroxamate resistance protein (argR). Analysis of this protein sequence reveals the following:

35 Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 40 bacterial cytoplasm --- Certainty=0.3252(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10269> which encodes amino acid sequence <SEQ ID 10270> was also identified.

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA88596 GB:M18729 unknown protein [Streptococcus pneumoniae]
 Identities = 63/141 (44%), Positives = 90/141 (63%)

50 Query: 4 MNKIERQKRIKRLIQSGQIGTQEIKHLKNEGIDVTQATLSRDLREIGLLKLSPEGKL 63
 M K +R + IK++I ++ TQ+EI+ L+ + VTQ TLSRDLREIGL K++ +
 Sbjct: 1 MRKRDRLHQLIKKMITTEKLSTQKEIQDRLEAHNVCVTQTTLRDLREIGLTKVKKNDMVY 60

Query: 64 YVSLSTATSNRFPALRSYILKVSASFMLVLTNLGEASVLANFIDEKGLPEILGTMA 123
 Y ++ L ++ V+RA F LVL+T LGEASVLAN +D ILGT+AG
 55 Sbjct: 61 YVLVNETEKIDLVEFLSHHLEGVAREFTLVLTHTKLGEASVLANIVDVNKDEWILGTVAG 120

Query: 124 ADTLLVICQNEIDIAKVFKEKL 144
 A+TLLVIC+++ +AK+ E L
 Sbjct: 121 ANTLLVICRDQHVAKLMEDRL 141

-1200-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3317> which encodes the amino acid sequence <SEQ ID 3318>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3176 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 101/145 (69%), Positives = 121/145 (82%)

Query: 4 MNKIERQKRIKRLIQSGQIGTQEEIKLHLKNEGIDVTQATLSRDLREIGLLKLRSPGK 63
 MNK+ERQ++IKR+IQ+ IGTQE+IK HL+ EGI VTQATLSRDLREIGLLKLR +GKL
 Sbjct: 1 MNKMERQQQIKRI IQAEHIGTQEDIKNHLQKEGIVVTQATLSRDLREIGLLKLRDEQ 60

Query: 64 YYSLSSTATSNRFSFALRSYILKVSASFMLVLNTNLGEASVLNFIDEKGLPEILGT 123
 YYSLS + FSP +R Y+LKV RA FMLVL+TNLGEA VLAN ID + +ILGT+AG
 Sbjct: 61 YYSLSFVATPFSPEVRFVYLKVDRAFGMLVLHTNLGEADVLANLIDNDAIEDILGTIAG 120

Query: 124 ADTLLVICQEDIAKVFEKELSVGL 148
 ADTLLVIC++E+IAK FEK+L+ GL
 Sbjct: 121 ADTLLVICRDEEIAKRFEKDLAAGL 145

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1077

A DNA sequence (GBSx1151) was identified in *S.agalactiae* <SEQ ID 3319> which encodes the amino acid sequence <SEQ ID 3320>. This protein is predicted to be DNA mismatch repair protein hexa (mutS). Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3570 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA88597 GB:M18729 mismatch repair protein [Streptococcus pneumoniae]
 Identities = 593/858 (69%), Positives = 698/858 (81%), Gaps = 14/858 (1%)

Query: 1 MAKPTISPGMQQYLDIKENYPDAFLFRMGDFYELFYDDAVKAAQILEISLTSRNKNAEK 60
 MA +SPGMQQY+DIK+ YPDAFLFRMGDFYELFY+DAV AAQILEISLTSRNKNA+
 Sbjct: 1 MAIEKLSPGMQQYVDIKQYPDAFLFRMGDFYELFYEDAVNAAQILEISLTSRNKNADN 60

Query: 61 PIPMAGVPYHSAQQYIDVLVELGYKVAIAEQMEDPKKAVGVVVKREVQVVTPTGTVVESTK 120
 PIPMAGVPYHSAQQYIDVL+E GYKVAIAEQMEDPK+AVGVVVKREVQV+TPGTVV+S+K
 Sbjct: 61 PIPMAGVPYHSAQQYIDVLIEQGYKVAIAEQMEDPKQAVGVVVKREVQVITPTGTVVDSSK 120

Query: 121 PDSANNFLVAIDSQDQQTFLAYMDVSTGEFQATLLTDFESVRSEILNLKAREIVVGYQL 180
 PDS NNFLV+ID + Q FGLAYMD+ TG+F T L DF V EI NLKARE+V+GY L
 Sbjct: 121 PDSQNNFLVSIIDREGNQ-FGLAYMDLVTGDFYVTGLLDFTLVCGEIRNLKAREVVLGYDL 179

Query: 181 TDEKNHLLTKQMNLLSYEDERLNDIHLIDEQLTDLEISAAEKLLQYVHRTQKRELSHLQ 240
 ++E+ +L++QMNLL+LSYE E D+HL+D +L +E +A+ KLLQYVHRTQ REL+HL+

-1201-

Sbjct: 180 S E E E E Q I L S R Q M N L V L S Y E K E S F E D L H L L D L R L A T V E Q T A S S K L L Q Y V H R T Q M R E L N H L K 239

Query: 241 K V V H Y E I K D Y L Q M S Y A T K N S L D L L E N A R T S K K H G S L Y W L L D E T K T A M G T R M L R T W I D R P L 300
V+ YEIKD+LQM YATK SLDL+ENAR+ KK GSL+WLLDETKTAMG R+LR+WI RPL

5 Sbjct: 240 F V I R Y E I K D F L Q M D Y A T K A S L D L V E N A R S G K K Q G S L F W L L D E T K T A M G M R L L R S W I H R P L 299

Query: 301 V S M N R I K E R Q D I I Q V F L D Y F F E R N D L T E S L K G V Y D I E R L A S R V S F G K A N P K D L L Q L G Q T L 360
+ RI +RQ+++QVFLD+FFER+DLT+SLKGVYDIERLASRVSGK NPKDLLQL TL

10 Sbjct: 300 I D K E R I V Q R Q E V V Q V F L D H F F E R S D L T D S L K G V Y D I E R L A S R V S F G K T N P K D L L Q L A T T L 359

Query: 361 S Q I P R I K M I L Q S F N Q P E L D I I V N K I D T M P E S L I N T A I A P E A Q A T I T E G N I I K S G F D K Q 420
S +PRI+ IL+ QP L ++ +D +PEESLI+ AIAPEA IT+G II++GFD+

15 Sbjct: 360 S S V P R I R A I L E G M E Q P T L A Y L I A Q L D A I P E E S L I S A A I A P E A P H V I T D G G I I R T G F D E T 419

Query: 421 L D N Y R T V M R E G T G W I A D I E A K E R A A S G I G T L K I D Y N K K D G Y Y F H V T N S N L S L V P E H F F R K 480
LD YR V+REGT WIA+IEAKER SGI TLKIDYNKKDGYFHV TNS L VP HFFRK

20 Sbjct: 420 L D K Y R C V L R E G T S W I A E I E A K E R E N S G I S T L K I D Y N K K D G Y Y F H V T N S Q L G N V P A H F F R K 479

Query: 481 A T L K N S E R Y G T A E L A K I E G E M L E A R E Q S S N L E Y D I F M R V R A Q V E S Y I K R L Q E L A K T I A T V 540
ATLKNSER+GT ELA+IEG+MLEARE+S+NLEY+IFMR+R +V YI+RLQ LA+ IATV

25 Sbjct: 480 A T L K N S E R F G T E E L A R I E G D M L E A R E K S A N L E Y E I F M R I R E E V G K Y I Q R L Q A L A Q G I A T V 539

Query: 541 D V L Q S L A V V A E N Y H Y V R P K F N D Q H Q I K I K N G R H A T V E K V M G V Q E Y I P N S I Y F D S Q T D I Q L 600
DVLQSLAVVAE H +RP+F D QI I+ GRHA VEKVMG Q YIPN+I T IQL

30 Sbjct: 540 D V L Q S L A V V A E T Q H L I R P E F G D S Q I D I R K G R H A V V E K V M G A Q T Y I P N T I Q M A E D T S I Q L 599

Query: 601 I T G P N M S G K S T Y M R Q L A L T V I M A Q M G G F V S A D E V D L P V F D A I F T R I G A A D D L I S G Q S T F M 660
+TGPNMSGKSTYMRQLA+T +MAQ+G +V A+ LP+FDAIFTRIGAADDL+SGQSTFM

35 Sbjct: 600 V T G P N M S G K S T Y M R Q L A M T A V M A Q L G S Y V P A E S A H L P I F D A I F T R I G A A D D L V S G Q S T F M 659

Query: 661 V E M M E A N Q A V K R A S D K S L I L F D E L G R G T A T Y D G M A L A Q S I I E Y I H D R V R A K T M F A T H Y H E 720
VEMMEAN A+ A+ SLILFDELGRGTATYDGMALAQSIIEYIH+ + AKT+FATHYHE

40 Sbjct: 660 V E M M E A N N A I S H A T K N S L I L F D E L G R G T A T Y D G M A L A Q S I I E Y I H E H I G A K T L F A T H Y H E 719

Query: 721 L T D L S E Q L T R L V N V H V A T L E R D G E V T F L H K I E S G P A D K S Y G I H V A K I A G L P I D L L D R A T D 780
LT L L L VNVHVATLE+DG+VTFLHKIE GPADKSYGIHVAKIAGLP DIL RA

45 Sbjct: 720 L T S L E S S L Q H L V N V H V A T L E Q D G Q V T F L H K I E G P A D K S Y G I H V A K I A G L P A D L L A R A D K 779

Query: 781 I L S Q L E A D A V Q L I V S P S Q E A V T A D L N E E L D S E K Q Q G Q L S L F E E P S N A G R V I E E L A I D I M 840
IL+QLE + SP T+ + E Q+SLF+ + ++ EL +D+

50 Sbjct: 780 I L T Q L E N Q G T E --- S P P P M R Q T S A V T E ----- Q I S L F D R - A E E H P I L A E L A K L D V Y 826

Query: 841 N L T P M Q A M N A I F D L K K L L 858
N+TPMQ MN + +LK+ L

45 Sbjct: 827 N M T P M Q V M N V L V E L K Q K L 844

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3321> which encodes the amino acid sequence <SEQ ID 3322>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.38 Transmembrane 532 - 548 (532 - 549)

----- Final Results -----
bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 661/858 (77%), Positives = 746/858 (86%), Gaps = 7/858 (0%)

60 Query: 1 M A K P T I S P G M Q Q Y L D I K E N Y P D A F L L F R M G D F Y E L F Y D D A V K A A Q I L E I S L T S R N K N A E K 60
M A K I S P G M Q Q Y L D I K + Y P D A F L L F R M G D F Y E L F Y + D A V K A A Q + L E I L T S R N K N A E
Sbjct: 1 M A K T N I S P G M Q Q Y L D I K D Y P D A F L L F R M G D F Y E L F Y E D A V K A A Q L L E I G L T S R N K N A E N 60

65 Query: 61 P I P M A G V P Y H S A Q Q Y I D V L V E L G Y K V A I A E Q M E D P K K A V G V V K R E V V Q V V T P G T V V E S T K 120

-1202-

PIPMAGVP+HSAQQYIDVL+ELGYKVA+AEQMEDPK+AVGVVKREVVQV+TPGTVV+S K
 Sbjct: 61 PIPMAGVPHHSAQQYIDVLIELGYKVAEAEQMEDPKQAVGVVKREVVQVITPGTVVDSAK 120

5 Query: 121 PDSANNFLVAIDSQDQQTFLGLAYMDVSTGEFQATLLTDFESVRSEILNLKAREIVVGYQL 180
 PDSANNFLVA+D D +GLAYMDVSTGEF T L DF SVRSEI NLKA+E+++G+ L
 Sbjct: 121 PDSANNFLVAVDL-DGCRYGLAYMDVSTGEFCVTDLADFTSVRSEIQNLKAKEVLLGFDL 179

10 Query: 181 TDEKNHLLTKQMNLLLSYEDERLNDIHLIDEQLTDLEISAAEKLLQYVHRTQKRELSHLQ 240
 ++E+ +L KQMNLLLSYE+ D LID QLT +E++AA KLLQYVH+TQ RELSHLQ
 Sbjct: 180 SEEEQTILVKQMNLLLSYEETVYEDKSLIDGQLTTVELTAAGKLLQYVHKTQMRLESHLQ 239

15 Query: 241 KVVHYEIKDYLMQSYATKNSLDLLENARTSKKHGSLYWLLDETCTAMGTRMLRTWIDRPL 300
 +VHYEIKDYLMQSYATK+SLDL+ENART+KKHGSYLWLLDETCTAMG R+LR+WIDRPL
 Sbjct: 240 ALVHYEIKDYLMQSYATKSSLDLVENARTNKKHGSYLWLLDETCTAMGMRLRSWIDRPL 299

20 Query: 301 VSMNRIKERQDIIQVFLDYFFERNDLTESLKGVDIERLASRVSFSGKANPKDLLQLGQTL 360
 VS I ERQ+IIQVFL+ F ER DL+ SLKGVDIERL+SRVSFGKANPKDLLQLG TL
 Sbjct: 300 VSKAAILERQEI IQVFLNAFIERTDLSNSLKGVDIERLSSRVSFSGKANPKDLLQLGHTL 359

25 Query: 421 LDNYRTVMREGTGWADIIEAKERASGIGTLKIDYNKKDGYFHVINSNLSLVPEHFFRK 480
 LD+YR VMREGTGWADIIEAKER ASGI LKIDYNKKDGYFHVINSNLSLVPEHFFRK
 Sbjct: 420 LDHYRKVMREGTGWADIIEAKERQASGINNLKIDYNKKDGYFHVINSNLSLVPEHFFRK 479

30 Query: 481 ATLKNSEYGTAEELAKIEGEMLEAREQSSNLEYDIFMRVRAQVESYIKRLQELAKTIATV 540
 ATLKNSEYGTAEELAKIEG+MLEARE+SS+LEYDIFM +RAQVE+YI RLQ+LAK +ATV
 Sbjct: 480 ATLKNSEYGTAEELAKIEGQMLEARESSSLEYDIFMCIRAQVETYINRLQKLAKILATV 539

35 Query: 541 DVLQSLAVVAENYHYVRPKFNDQHQIKIKNGRHATVEKVMGVQEIYIPNSIYFDSQTDIQL 600
 DVLQSLAVVAE HY+RP+FND H I I+ GRHA VEKVMGVQEIYIPNSI FD QT IQL
 Sbjct: 540 DVLQSLAVVAETNHYTRPQFNDNHVITIQEGRHAVVEKVMGVQEIYIPNSISFDQQTSLQL 599

40 Query: 601 ITGPNMSGKSTYMRQLALTIVMAQMGGFVSADEVLDLPVFDALFTRIGAADDLISGQSTFM 660
 ITGPNMSGKSTYMRQLALTIVMAQMG FV+AD VDLF+FDALFTRIGAADDLISGQSTFM
 Sbjct: 600 ITGPNMSGKSTYMRQLALTIVMAQMGSFVAADHVDLPFLFDALFTRIGAADDLISGQSTFM 659

45 Query: 661 VEMMEANQAVKRASDKSLILFDELGRGTATYDGMALAQSIIEYIHDRVRAKTMFATHYHE 720
 VEMMEANQA+KRASD SLILFDELGRGTATYDGMALAQ+IIEYIHDRV AKT+FATHYHE
 Sbjct: 660 VEMMEANQAIKRASDNSLILFDELGRGTATYDGMALAQAIIEYIHDRVGAKTIFATHYHE 719

50 Query: 721 LTDLSEQLTRLNVNHVATLERDGEVTFHLKIESGPADKSYGIHVAKIAGLPIDLLDRATD 780
 LTDLS LT LVNVHVATLE+DG+VTFHLKI GPADKSYGIHVAKIAGLP LL RA +
 Sbjct: 720 LTDLSLNLTSLVNVHVATLEKDGVDVTFHLKIAEGPADKSYGIHVAKIAGLPKSLKRADE 779

55 Query: 781 ILSQLEADAVQLIVSPSQEAVTADLNEELDSEKQGGQLSLFEEPSNAGRVIEELEAIDIM 840
 +L++LE S S E ++ E S +QGQLSLF + A + + LE ID+M
 Sbjct: 780 VLTRLETQ-----SRSTEIISVPSQVESSSAVRQGQLSLFGDEEKAHEIRQALEVIDVM 833

Query: 841 NLTPMQAMNAIFDLKLL 858
 N+TP+QAM +++LKKLL
 Sbjct: 834 NMTPLQAMTTLVELKLL 851

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1078

A DNA sequence (GBSx1152) was identified in *S.agalactiae* <SEQ ID 3323> which encodes the amino acid sequence <SEQ ID 3324>. This protein is predicted to be cold shock protein-related protein. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

-1203-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2095 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB69404 GB:A91080 unnamed protein product [unidentified]
 Identities = 48/63 (76%), Positives = 56/63 (88%)

10

Query: 1 MTQGTVKWPNSEKGFISSETGTDVFAHFSEIKVDGFKTLEEGQKVTFDIQDGGQGPQA 60
 MT+GTVKWFN +KGFSGFI+SE G DVFAHFS+I+ GFKTL+EGQKVTFD++ GQGPQA
 Sbjct: 1 MTKGTVKWFNPDKGFGFITSEDGQDVFAHFSQIQTSFGFKTLDEGQKVTFDVEAGQGPQA 60

15

Query: 61 TNI 63
 NI
 Sbjct: 61 VNI 63

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3325> which encodes the amino acid
 sequence <SEQ ID 3326>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25

 bacterial cytoplasm --- Certainty=0.2350 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30

Identities = 49/63 (77%), Positives = 56/63 (88%)

Query: 1 MTQGTVKWPNSEKGFISSETGTDVFAHFSEIKVDGFKTLEEGQKVTFDIQDGGQGPQA 60
 M QGTVKWFN+EKGFGFIS+E G DVFAHFS I+ +GFKTLEEGQKV FD+++GQGPQA
 Sbjct: 3 MAQGTVKWFNAEKGFISTENGQDVFAHFSAIQTNGFKTLEEGQKVAFDVEEGQGPQA 62

35

Query: 61 TNI 63
 NI
 Sbjct: 63 VNI 65

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1079

A DNA sequence (GBSx1153) was identified in *S.agalactiae* <SEQ ID 3327> which encodes the amino
 acid sequence <SEQ ID 3328>. Analysis of this protein sequence reveals the following:

45

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50

 bacterial cytoplasm --- Certainty=0.6378 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1080

A DNA sequence (GBSx1154) was identified in *S. agalactiae* <SEQ ID 3329> which encodes the amino acid sequence <SEQ ID 3330>. This protein is predicted to be DNA mismatch repair protein hexb (mutL). Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2242(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10267> which encodes amino acid sequence <SEQ ID 10268> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA88600 GB:M29686 mismatch repair protein [Streptococcus pneumoniae]
Identities = 452/657 (68%), Positives = 543/657 (81%), Gaps = 8/657 (1%)

Query: 20 LSKIIELPDILANQIAAGEVVERPSSVVKELVENAIDAGSSQITIEVEESGLKKIQITDN 79
+S IIBLP++LANQIAAGEV+ERP+SV KELVENAIDAGSSQI IE+EE+GLKK+QITDN
Sbjct: 1 MSHIIEPEMLANQIAAGEVIERPASVCKELVENAIDAGSSQIIIEIEEAGLKKVQITDN 60

Query: 80 GEGMTSEDAVLSLRRHATSKIKSQSDLFRIRTLGFRGEALPSIASISLMTIKTATEQKGK 139
G G+ ++ L+LRRHATSKIK+Q+DLFRIRTLGFRGEALPSIAS+S++T+ TA +
Sbjct: 61 GHGIAHDEVELALRRHATSKIKNQADLFRIRTLGFRGEALPSIASVSVLTLLTAVDGDASH 120

Query: 140 GTLLVAKGGNIEKQEVVSSPRGTKILVENLFFNTPARLKYMKSLQSELAHIIDIVNRLSL 199
GT LVA+GG +E+ +SP GTK+ VE+LFFNTPARLKYMK S+EL+HIIDIVNRL L
Sbjct: 121 GTKLVARGGEVEEVIPATSPVGTKVCVEDLFFNTPARLKYMKSQQAELSHIIDIVNRLGL 180

Query: 200 AHPEVAFTLINDGKEMTKTSGTDLRQAIAGIYGLNTAKKMI EISNADLD FEISGYVSLP 259
AHPE++F+LI+DGKEMT+T+GTG LRQAIAGIYGL +AKKMI EI N+DLDFEISG+VSLP
Sbjct: 181 AHPEISFSLISDGKEMTRTAGTGQLRQAIAGIYGLVSAKKMIETENSDDLDFEISGFVSLP 240

Query: 260 ELTRANRNYITLLINGRYIKNFFLNRSILDGYGSKLMVGRFPPIAVIDIQIDPYLADVNVH 319
ELTRANRNYI+L INGRYIKNFFLN+ILDG+GSKLMVGRFP+AVI I IDPYLADVNVH
Sbjct: 241 ELTRANRNYISLFINGRYIKNFFLNRAILDGFGSKLMVGRFPLAVIHIHIDPYLADVNVH 300

Query: 320 PTKQEVRIKSKERELMSLISTAISSESLKQYDLIPDALENLAKTSTRSVDPKPIQTSFSLKQP 379
PTKQEVRIKSK+ELM+L+S AI+ SLK+ LIPDALENLAK++ R+ +K QT LK+
Sbjct: 301 PTKQEVRIKSKELMTLVSEAIANSLSKEQTLIPDALENLAKSTVRNREKVEQTILPLKEN 360

Query: 380 GLYYDRAKNDFFIGADTVSEPIANFTNLDKSDGSDNDVKNVSNQGATQSPNIKYASRDQ 439
LYY++ + + +E L + K ++++ T+ + +A R
Sbjct: 361 TLYYEKTEP----SRPSQTEVADYQVELTDEGQDLTLFAKETLDR-LTKPAKLHFAERKP 415

Query: 440 ADSENFHISQDYLSKQSLNKLVEKLDSEESSTFPELEFFFGQMHGTYLFAQGNGGLYIID 499
A+ + H + L+ S++K +KL+ EE+S+FPELEFFFGQMHGTYLFAQG GLYIID
Sbjct: 416 ANYDQLDHPDLA---SIDKAYDKLEREEASSFPELEFFFGQMHGTYLFAQGRDGLYIID 472

Query: 500 QHAAQERVVKY EYREKIGEVDNSLQQLLVFPFLFEFSSDFLQLEKMSLLQDVGFIFLEPY 559
QHAAQERVVKY YRE IG VD S QQLLV++FEF + D L+L+E+M LL++VG+FL Y
Sbjct: 473 QHAAQERVVKY EYRESIGNVDQSQQQLLVFPIFEFPADDA RLKERMPLLEE VGVFLAEY 532

Query: 560 GNNTFILREHPIWMKEEVEESGIYEMCDMLLLTNEVSVKKYRAELAIMMSCKRSIKANHT 619
G N FILREHPIWM EEE+ESGIYEMCDMLLLT EVS+KKYRAELAIMMSCKRSIKANH
Sbjct: 533 GENQFILREHPIWMAEEBIESGIYEMCDMLLLTKEVSIKKYRAELAIMMSCKRSIKANHR 592

-1205-

Query: 620 LDDYSARHLDDQLAQCKNPYNCPHGRPVLVNFVKADMEKMFRIQENHTSLRDLGKY 676
 +DD+SAR LL QL+QC NPYNCPHGRPVLV+FTK+DMEKMF+RIQENHTSLR+LGKY
 Sbjct: 593 IDHSARQLLYQLSQCDNPYNCPHGRPVLVHFTKSDMEKMFRIQENHTSLRELKY 649

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3331> which encodes the amino acid sequence <SEQ ID 3332>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1854(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 502/663 (75%), Positives = 574/663 (85%), Gaps = 9/663 (1%)

- Query: 20 LSKIIELPDILANQIAAGEVVERPSSVVKELVENAIDAGSSQITIEVEESGLKKIQITDN 79
 ++ IIELP++LANQIAAGEVVERP+SVVKELVENAIDA SSQIT+E+EESGLK IQ+TDN
 20 Sbjct: 14 MTNIIELPEVLANQIAAGEVVERPASVVKELVENAIDAKSSQITVEIEESGLKMIQVTDN 73
- Query: 80 GEGMTSEDAVLSLRRHATSKIKSQSDLFRIRTLGFGEALPSIASISLMTIKTATEQGGK 139
 GEGM+ ED LSLRRHATSKIKSQSDLFRIRTLGFGEALPS+ASIS +TIKTAT++
 25 Sbjct: 74 GEGMSHEDLPLSLRRHATSKIKSQSDLFRIRTLGFGEALPSVASISKITIKTATKEVTH 133
- Query: 140 GTLLVAKGGNIEKQEVVSSPRGKILVENLFFNTPARLKYMKSLSQELAHIIDIVNRLSL 199
 G+LL+A GG IE E +S+P GTKI VENLF+NTPARLKYMKSLSQ+ELAHID+VNRLSL
 30 Sbjct: 134 GSLLIATGGEIETLEAISTPTGTIKIKVENLFFNTPARLKYMKSLSQELAHIVDVVNRLSL 193
- Query: 200 AHPEVAFTLINDGKEMTKTSGTCDLRQAIAGIYGLNTAKKMIIEISNADLDFEISGYVSLP 259
 AHPEVAFTLI+DG+++T+TSGTCDLRQAIAGIYGLNT KKM+ ISNADLDFE+SGYVSLP
 35 Sbjct: 194 AHPEVAFTLISDGRQLTQTSGTCDLRQAIAGIYGLNTTKMLAISNADLDFEVSQYVSLP 253
- Query: 260 ELTRANRNYITLLINGRYIKNFFLNRSILDGYGSKLMVGRFPIAVIDIQIDPYLADVNVH 319
 ELTRANRNY+T+L+NGRYIKNFFLN+ILDGYGSKLMVGRFPI VIDIQIDPYLADVNVH
 40 Sbjct: 254 ELTRANRNYMTILVNGRYIKNFFLNRAILDGYGSKLMVGRFPIVVIDIQIDPYLADVNVH 313
- Query: 320 PTKQEVRIKSKERELMSLISTAISESLKQYDLIPDALENLAKTSTRSVDKPIQTSFSLKQP 379
 PTKQEVRIKSKERELM+LISTAISESLK+ DLIPDALENLAK+STR KP QT L+
 45 Sbjct: 314 PTKQEVRIKSKERELMALISTAISESLKEQDLIPDALENLAKSSTRHFSKPEQTQLPLQSR 373
- Query: 380 GLYYDRAKNDFFIGADTVSEPIANFTNLDKSDGSDNDVNKNSV-----NQGATQSPNIK 433
 GLYYD KNDFF+ VSE I D G+VDN VK ++ ++K
 50 Sbjct: 374 GLYYDPQKNDFFVKESAVSEKI---PFTDFYSGAVDNSVKVEKVELLPHSEEVIGPSSVK 430
- Query: 434 YASRDQADSENFHISQDYLSSQSLNKLVEKLDSEESSTFPELEFFGQMHGTYLFAQGNG 493
 +ASR Q H L ++Q L++++ +L++E S FPEL++FGQMHGTYLFAQG
 55 Sbjct: 431 HASRPQNTFTETDHPNLDLKNRQKLSQMLTRLENEGQSVFPEDLYFGQMHGTYLFAQGD 490
- Query: 494 GLYIIDQHAQAQERVKYEYRREKIGEVDNSLQQLLVPFLFEFSSSDFLQLEKMSLLQDVG 553
 GL+IIDQHAQAQERVKYEYR+KIGEVD+SLQQLLVP+LFEFS SDF+ LQEKM+LL +VG
 60 Sbjct: 491 GLFIIDQHAQAQERVKYEYRDKIGEVDSSLQQLLVPYLFEFSGSDFINLQEKMALLNEVG 550
- Query: 554 IFLEPYGNNTFILREHPIMWKKEEVEESGIYEMCDMLLLTNEVSVKKYRAELAIMMSCKRS 613
 IFLE YG+NTFILREHPIMWKKEE+ SG+YEMCDMLLLTNEVS+K YRAELAIMMSCKRS
 55 Sbjct: 551 IFLEVYGHNTFILREHPIMWKKEEIASGVYEMCDMLLLTNEVSIKTYRAELAIMMSCKRS 610
- Query: 614 IKANHLLDDYSARHLDDQLAQCKNPYNCPHGRPVLVNFVKADMEKMFRIQENHTSLRDLGKY 676
 IKANH+LDDYSAR+LL QLAQC+NPYNCPHGRPVL+NF+KADMEKMF+RIQENHTSLR+LGKY
 60 Sbjct: 611 IKANHSLLDDYSARNLLQLAQCNPNYNCPHGRPVLINFSKADMEKMFRIQENHTSLRELKY 673

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1206-

Example 1081

A DNA sequence (GBSx1155) was identified in *S.agalactiae* <SEQ ID 3333> which encodes the amino acid sequence <SEQ ID 3334>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3372(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
15 vaccines or diagnostics.

Example 1082

A DNA sequence (GBSx1156) was identified in *S.agalactiae* <SEQ ID 3335> which encodes the amino acid sequence <SEQ ID 3336>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
20 >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -14.01    Transmembrane 176 - 192 ( 170 - 197)
      INTEGRAL    Likelihood = -8.07     Transmembrane 390 - 406 ( 387 - 412)
      INTEGRAL    Likelihood = -6.10     Transmembrane 271 - 287 ( 269 - 291)
25     INTEGRAL    Likelihood = -6.00     Transmembrane 83 - 99 ( 82 - 101)
      INTEGRAL    Likelihood = -4.78     Transmembrane 51 - 67 ( 50 - 71)
      INTEGRAL    Likelihood = -2.92     Transmembrane 303 - 319 ( 302 - 320)
      INTEGRAL    Likelihood = -2.76     Transmembrane 363 - 379 ( 362 - 381)
      INTEGRAL    Likelihood = -2.39     Transmembrane 152 - 168 ( 151 - 169)
      INTEGRAL    Likelihood = -2.02     Transmembrane 325 - 341 ( 325 - 342)
30     INTEGRAL    Likelihood = -1.65     Transmembrane 226 - 242 ( 226 - 242)
      INTEGRAL    Likelihood = -0.90     Transmembrane 24 - 40 ( 24 - 40)
      INTEGRAL    Likelihood = -0.27     Transmembrane 111 - 127 ( 111 - 127)

----- Final Results -----
35     bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10265> which encodes amino acid sequence <SEQ ID
40 10266> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA61918 GB:X89779 LmrP integral membrane protein [Lactococcus
      lactis]
      Identities = 145/401 (36%), Positives = 236/401 (58%), Gaps = 4/401 (0%)
45
Query: 9  VKEFFALPKQLQLRELLRFISITVGS AIFPFMAMYYVQYFGNLVTGILIIITQLSGFVAT 68
      +KEF+ L K LQLR + F+ +F M +YY QY G+ +IGIL+ ++ ++ FVA
Sbjct: 1  MKEFWNLDRNLQLRLGIVFLGAFSYGTVFSSMTIYYNQYLGSAITGILLALS AVATFVAG 60

Query: 69 LYGGHLS DAMGRKKVVIIGSLLATIGWAITIAANVPNHITPHLTFVGILIIIEIAHQFYFP 128
      + G +D GRK V++ G+++ +G A+ IA+N+P H+ P TF+ L+I + F
Sbjct: 61 ILAGFFADRNGRKPVMVFGTIIQLLGAALAIASNLPGHVNPWSTFIAPLLISFGYNFVIT 120

Query: 129 AYEAMTIDLTNEQNRRFVYTIGYWLNVNIAVMLGSGIAGIFYDHHFFELLIVLLIISAICC 188
55     A AM ID +N +NR+ V+ + YW N++V+LG+ + + F LL++LL+ +

```

-1207-

5 Sbjct: 121 AGNAMIIDASNAENRKVVFM LDYWAQNLSVILGAALGAWLFRPAFEALLVILLTTLVLSF 180
 Query: 189 FVVYFKFDET-KPQEGTFKHDKGVLGTFKNYSQVLVDKAFVVYTLGAIGSSVVWLQVDNY 247
 F+ F ET KP T K D+ F+ Y VL DK ++++ I ++ + +Q DN+
 10 Sbjct: 181 FLTTFVMTETFKP---TVKVDEKAENIFQAYKTVLQDKTYMIFMGANIATTFIIMQFDNF 237
 Query: 248 FSVNLKQNFVVSILGHTITGAKMLSLAVFTNTLLIVLLMTTINKFIENWPLKRLILGS 307
 V+L +F+ ++ G I G +ML++ + +L+VLLMTT+N+ ++W ++ I GS
 15 Sbjct: 238 LPVHLSNSFKTITFWGFEIYGQRM LTIYLILACVLVLLMTTLNRLTKDWSHQKGFIVGS 297
 Query: 308 LICGFGMLFNISLNTFGAILIAMTFFTFGEMIVPASQVLRRAEMMVEGKIGSYSGFLAIA 367
 L GM+F+ TF I IA +T GE++Y P+ Q L A++M KIGSY+G AI
 Sbjct: 298 LFMAIGMIFSLTTTTFPIFIAGIVYTLGEIVYTPSVQTLGADLMNPEKIGSYNGVAAIK 357
 20 Query: 368 QPVASVLAGAMVSLSYFTGKIGVQITLTIFMLAGLVILLYA 408
 P+AS+LAG +VS+S IGV + L + + ++L+L A
 Sbjct: 358 MPIASILAGLLVSI SPMIKAIGVSLVLALTEVLAIILVLVA 398

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3337> which encodes the amino acid
 sequence <SEQ ID 3338>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 25 INTEGRAL Likelihood = -11.41 Transmembrane 166 - 182 (161 - 188)
 INTEGRAL Likelihood = -7.75 Transmembrane 384 - 400 (376 - 403)
 INTEGRAL Likelihood = -7.64 Transmembrane 266 - 282 (261 - 285)
 INTEGRAL Likelihood = -4.25 Transmembrane 295 - 311 (291 - 313)
 INTEGRAL Likelihood = -2.71 Transmembrane 98 - 114 (98 - 115)
 INTEGRAL Likelihood = -2.23 Transmembrane 355 - 371 (355 - 374)
 30 INTEGRAL Likelihood = -2.02 Transmembrane 218 - 234 (218 - 234)
 INTEGRAL Likelihood = -1.91 Transmembrane 315 - 331 (315 - 331)
 INTEGRAL Likelihood = -1.22 Transmembrane 75 - 91 (75 - 92)
 INTEGRAL Likelihood = -0.75 Transmembrane 45 - 61 (45 - 63)
 INTEGRAL Likelihood = -0.75 Transmembrane 144 - 160 (144 - 161)
 35 ----- Final Results -----
 bacterial membrane --- Certainty=0.5564(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the databases:

>GP:CAA61918 GB:X89779 LmrP integral membrane protein [Lactococcus
 lactis]
 Identities = 138/400 (34%), Positives = 223/400 (55%), Gaps = 2/400 (0%)
 45 Query: 1 MQEFNLPRKQIQRLQRLVRFVITITLGSSIFFMAMYYTTYFGTFWIGLLMMITSLMGFVGT 60
 M+EF NL K +QLR + F+ ++F M +YY Y G+ TG+L+ ++++ FV
 Sbjct: 1 MKEFWNLDRKQLRLGIVFLGAFSYGTVFSSMTIYYNQYLGSAITGILLALSAVATFVAG 60
 50 Query: 61 LYGGHLSDALGRKKVIMIGSVGTTLGWFLTILANLPNAAIPWLTFAGILIVEIASSFYGP 120
 + G +D GRK V++ G++ LG L I +NLP PW TF LL+ +F
 Sbjct: 61 ILAGFFADRNGRKPVMVFGTIIQLLGAALALASNLPGHVNPWSTFIAFLILISFGYNFVIT 120
 Query: 121 AYEAMLIDLITDESNNRFFVYTYNYWFINIAVMFGAGLSGLFYDHHFLALLVALLLVNVLCF 180
 A AM+ID ++ NR+ V+ ++YW N++V+ GA L + F ALLV LLL ++ F
 55 Sbjct: 121 AGNAMIIDASNAENRKVVFM LDYWAQNLSVILGAALGAWLFRPAFEALLVILLTTLVLSF 180
 Query: 181 GVAYYCFDETRPETHAFDHGKLLASFNQYRQVFDRAFLVFTLGAIFSGSIWMQMDNYV 240
 + + ET T D + FQ Y+ V D+ ++F I + I MQ DN++
 60 Sbjct: 181 FLTTFVMTETFKPTVKVDEKAENI--FQAYKTVLQDKTYMIFMGANIATTFIIMQFDNFL 238
 Query: 241 PVHLKLYFQPTAVLGFQVTSKMSLMVLINTLLIVLFMTVVNKLTEKWKLPLQVLVVGSL 300
 PVHL F+ GF++ +ML++ ++ +L+VL MT +N+LT+ W + GSL
 Sbjct: 239 PVHLSNSFKTITFWGFEIYGQRM LTIYLILACVLVLLMTTLNRLTKDWSHQKGFIVGSL 298
 65 Query: 301 LFTLGMLLSFTFTQFYAIWLSVLLTTFGEMINVSASQVLRADMDHSQIGSYTGFSVMAQ 360

-1208-

+GM+ SF T F I+++ ++ T GE++ + Q L AD+M+ +IGSY G ++
 Sbjct: 299 FMAIGMIFSLTTFITPIFIAGIVYTLGEIVYTPSVQTLGADLMNPEKIGSYNGVAAIKM 358

Query: 361 PLGAILASLLVSVSHFTGPLGVQCLFAVIALLLGIYFTVVS 400
 P+ +ILA LLVS+S +GV + A+ +L I +V+
 Sbjct: 359 PIASILAGLLVSIISPMIKAIGVSLVLALTEVLAILVLVA 398

An alignment of the GAS and GBS proteins is shown below.

Identities = 228/406 (56%), Positives = 305/406 (74%)

Query: 9 VKEFFALPKQLQIRELLRFISITVGSALFFPMAMYYVQYFGNLVTCILIIITQLSGFVAT 68
 ++EF LPKQ+QLR+L+RF++IT+GS+IFFPMAMYY YFG TC+L++IT L GFV T
 Sbjct: 1 MQEFLNLPKQIQRLQRLVRFVITITLGSIIFFPMAMYYTTYFGTFWTGLLMMITSLMGFVGT 60

Query: 69 LYGGHLSADAMGRKKVVIIGSLATIGWAITIAANVPNHITPHLTFVGILIIETIAHQFYFP 128
 LYGGHLSDA+GRKKV++IGS+ T+GW +TI AN+PN P LTF GIL++EIA FY P
 Sbjct: 61 LYGGHLSDALGRKKVIMIGSVGTTLGGWFLTILANLPNAAIPWLTFAGILLVEIASSFYGP 120

Query: 129 AYEAMTIDLITNEQNRFFVYTIGYWLNVIAVMLGSGIAGIFYDHHFFELLIVLLIISAICC 188
 AYEAM IDLT+E NRRFVYTI YW +NIAVM G+G++G+FYDHHF LL+ LL+++ +C
 Sbjct: 121 AYEAMLIDLITDESNRFFVYTINYWFINIAVMFGAGLSGLFYDHHFLALLVALLLVNVLCP 180

Query: 189 FVVYFKFDETKPQEGTFKHDKGVLTGTFKNYSQVLVDKAFVYVYTLGAIGSSVVLQVDNYF 248
 V Y+ FDET+P+ F H KG+L +F+NY QV D+AFV++TLGAI S +W+Q+DNY
 Sbjct: 181 GVAYYCFDETRPETHAFDHGKGLLASFQNYRQVFHDRAFLVFTLGAIFSGSIWMQMDNYV 240

Query: 249 SVNLRKQNFVVSILGHTITGAKMLSLAVFTNTLLIVLLMTTINKFIENWPLKRQLILGSL 308
 V+LK F+ ++LG +T +KMLSL V TNTLLIVL MT +NK E W L QL++GSL
 Sbjct: 241 PVHLKLYFQPTAVLGFQVTSKMLSLMVLNTNTLLIVLFMTVVNKLTEKWKLLPQLVVGSL 300

Query: 309 ICGFGMLFNISLNTFGAILIAMTFFTFGEMIYVPASQVLRAEMMVEGKIGSYSGFLAIAQ 368
 + GML + + F AI +++ TFGEMI V ASQVLRA+MM +IGSY+GF+++AQ
 Sbjct: 301 LFTLGMLLSFTFTQFYAIWLSVLLTTFGEMINVSASQVLRAEDMDHSQIGSYTGFVSMQAQ 360

Query: 369 PVASVLGAMVLSYFTGKIGVQITLTIFFMLAGLVLLIYATKMKNI 414
 P+ ++LA +VS+S+FTG +GVQ + L G+ + + KMK +
 Sbjct: 361 PLGAILASLLVSVSHFTGPLGVQCLFAVIALLLGIYFTVVSARKMKV 406

A related GBS gene <SEQ ID 8725> and protein <SEQ ID 8726> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
 SRCFLG: 0
 McG: Length of UR: 4
 Peak Value of UR: 1.73
 Net Charge of CR: 1
 McG: Discrim Score: -4.26
 GvH: Signal Score (-7.5): -2.48
 Possible site: 35

>>> Seems to have no N-terminal signal sequence

Amino Acid Composition: calculated from 1

ALOM program count: 12 value: -14.01 threshold: 0.0

INTEGRAL	Likelihood = -14.01	Transmembrane	168 - 184 (162 - 189)
INTEGRAL	Likelihood = -8.07	Transmembrane	382 - 398 (379 - 404)
INTEGRAL	Likelihood = -6.10	Transmembrane	263 - 279 (261 - 283)
INTEGRAL	Likelihood = -6.00	Transmembrane	75 - 91 (74 - 93)
INTEGRAL	Likelihood = -4.78	Transmembrane	43 - 59 (42 - 63)
INTEGRAL	Likelihood = -2.92	Transmembrane	295 - 311 (294 - 312)
INTEGRAL	Likelihood = -2.76	Transmembrane	355 - 371 (354 - 373)
INTEGRAL	Likelihood = -2.39	Transmembrane	144 - 160 (143 - 161)
INTEGRAL	Likelihood = -2.02	Transmembrane	317 - 333 (317 - 334)
INTEGRAL	Likelihood = -1.65	Transmembrane	218 - 234 (218 - 234)
INTEGRAL	Likelihood = -0.90	Transmembrane	16 - 32 (16 - 32)
INTEGRAL	Likelihood = -0.27	Transmembrane	103 - 119 (103 - 119)
PERIPHERAL	Likelihood = 9.44		239

modified ALOM score: 3.30

*** Reasoning Step: 3

```

5      ----- Final Results -----
      bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

ORF01675(325 - 1530 of 1854)
EGAD|40187|42348(1 - 400 of 408) integral membrane protein (LmrP) {Lactococcus lactis}
GP|1052754|emb|CAA61918.1||X89779 LmrP integral membrane protein {Lactococcus lactis}
PIR|S58131|S58131 integral membrane protein LmrP - Lactococcus lactis
%Match = 21.7
%Identity = 36.2 %Similarity = 60.8
Matches = 145 Mismatches = 155 Conservative Sub.s = 99

[illegible]

25 483 513 543 573 603 633 663 693

VTGILIIITQLSGFVATLYGCHLSDAMGRKKVVIIGSLLATIGWAITIAANVPNHITPHLTFVGGILIIIEIAHQFYFPAYE
:||||: :: :: ||| : | ::| ||| |:: |::: :| ::| :|:| | :| :| :| ||: |:| :| |

ITGILLALSAVATFVAGILAGFADNRGRKPFVMVFGTIIQLLGAALAIASNLPGHVNPWSTFIAFLILISFGYNFVITAGN

 60 70 80 90 100 110 120

```

723      753      783      813      843      873      900      930
AMTIDLNEQNRRFVYITGYLWNIAVMLGSGIAGIFYDHHFFELLIVLLIISAICCFVVPYKFDET-KPQEGTFKHKDG
|| || : ||| : : : || |::||| : : : |||::|| : : | ||| || | | :
AMIIDASNAENRKVFMLDYWAQNLSVILGAALGAWLFRPAFEALLVILLTLVLSFFLTTFTVMETFKP---TVKVDEK
140      150      160      170      180      190      200

```

45

	1200	1230	1260	1290	1320	1350	1380	1410
	NKFIENWPLKRQLILGLSLICGFMLEFNISLNTFGAILIAMTFTTFGEMIVYPASQVLRAEMMVEGKIGSXSGLFLAIAQPV							
	:: :: :: : ::: : : ::: : :: : :							
	NRLTKDWSHQKGFIWGSLEFMAIGMIFSLTTFTPIFIAGIIVYTLGEIVYTPSVQTILGADLMNPEKIGSYNGVAAIKMPI							
	290	300	310	320	330	340	350	360

50 1440 1470 1500 1530 1560 1590 1620 1650
ASVLGAMVSLSYFTGKIGVQITLTIFMLAGLVLLIYATKMKNIEIGK*NVRLY*RKIE*NNG*IYCCGNSWIGIHDI
||:| | :| :| | | : | : : :| :| |
ASILAGLLVSI³⁷⁰SPMIKAIGVSLV³⁸⁰LALTEVL³⁹⁰LAII⁴⁰⁰LVAVNRHQKTKLN

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1083

A DNA sequence (GBSx1157) was identified in *S.agalactiae* <SEQ ID 3339> which encodes the amino acid sequence <SEQ ID 3340>. This protein is predicted to be holliday junction DNA helicase (ruvA).

60 Analysis of this protein sequence reveals the following:

Possible site: 37

-1210-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.75 Transmembrane 75 - 91 (74 - 91)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:BAB04943 GB:AP001511 holliday junction DNA helicase [Bacillus halodurans]
 Identities = 86/201 (42%), Positives = 122/201 (59%), Gaps = 6/201 (2%)

Query: 1 MYDYIKGKLSKITAKFIVVETAGLGYMIYVANPYSFSGYVNQEVITYLHQVIRDDAHLF 60
 M DY++G L+ I ++ VVE G+GY +Y NPY F + +TIY Q +R+D L+
 15 Sbjct: 1 MIDYLRGTLTDIDHQYAVVEVHGVGYQVYCPNPYEFKERDSVITYTFQYVREDVIRLY 60
 Query: 61 GFHTENEKEIFLNLISVSGIGPTTALAIIVDDNEGLVSAIDNSDIKYLTKFPKIGKKT 120
 GF T+ ++ +F L++VSGIGP ALAI+A E ++ AI+ D +L KFP +GKKT
 Sbjct: 61 GFRTKEKRSLFEKLLNVSGIGPKGALAILATGQPEHVIQAIEEDEAFVVKFPGVGKKT 120
 20 Query: 121 QQMILDLGKGFVE-----ASGESATSRKVSSEQNSNLEEAMEALLALGYKATELKVK 174
 +Q+ILDL GK E + E ++ N L+EAMEAL ALGY ELKKVK
 Sbjct: 121 RQIILDLGKGVDELHPGLFSQKEEQPKPHEKNDGNQALDEAMEALKALGYVEKELKKVKP 180

25 Query: 175 FFEGTNETVEQYIKSSLKML 195
 E T + YIK +L+++
 Sbjct: 181 KLEQETLTDDAYIKKALQLML 201

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3341> which encodes the amino acid
 30 sequence <SEQ ID 3342>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.59 Transmembrane 75 - 91 (74 - 91)

35 ----- Final Results -----

bacterial membrane --- Certainty=0.1638(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the databases:

>GP:BAB04943 GB:AP001511 holliday junction DNA helicase [Bacillus halodurans]
 Identities = 91/201 (45%), Positives = 128/201 (63%), Gaps = 5/201 (2%)

45 Query: 1 MYDYIKGQLTKITAKYIVVEANGLGYMINVANPYSFTDSVNQLVTIYLHQVIREDAHLF 60
 M DY++G LT I +Y VVE +G+GY + NPY F + ++TIY Q +RED L+
 Sbjct: 1 MIDYLRGTLTDIDHQYAVVEVHGVGYQVYCPNPYEFKERDSVITYTFQYVREDVIRLY 60
 Query: 61 GFHTEDEKDVFLKLISVSGIGPTTALAIIVDDNEGLVNAIDNSDIKYLKFPKIGKKT 120
 GF T++++ +F KL++VSGIGP ALAI+A E ++ AI+ D +L+KFP +GKKT
 50 Sbjct: 61 GFRTKEKRSLFEKLLNVSGIGPKGALAILATGQPEHVIQAIEEDEAFVVKFPGVGKKT 120
 Query: 121 QQMVLDLGKGFVEA-----POETGHTKARSNKAGNTQLDEAIEALLALGYKAKELKKIRA 175
 +Q++LDL GK E Q+ K GN LDEA+EAL ALGY KELKK++
 Sbjct: 121 RQIILDLGKGVDELHPGLFSQKEEQPKPHEKNDGNQALDEAMEALKALGYVEKELKKVKP 180
 55 Query: 176 FFEGTSETAQYIKSALKLLM 196
 E + T + YIK AL+L++
 Sbjct: 181 KLEQETLTDDAYIKKALQLML 201

60 An alignment of the GAS and GBS proteins is shown below.

Identities = 153/197 (77%), Positives = 176/197 (88%), Gaps = 1/197 (0%)

Query: 1 MYDYIKGKLSKITAKFIVVETAGLGYMIYVANPYSFSGYVNQEVITYLHQVIRDDAHLF 60

-1211-

MYDYIKG+L+KITAK+IVVE GLGYMI VANPYSF+ VNQ VTIYLHQVIR+DAHLLF
 Sbjct: 1 MYDYIKGQLTKITAKYIVVEANGLGYMINVANPYSFTDSVNQLVTIYLHQVIREDAHLLF 60
 Query: 61 GFHTENEKEIFLNLISVSGIGPTTALAI+AVDDNEGLVSAIDNSDIKYLTKFPKIGKKT 120
 5 GFHTE+EK++FL LISVSGIGPTTALAI+AVDDNEGLV+AIIDNSDIKYL KFPKIGKKT
 Sbjct: 61 GFHTEDEKDVFLKLISVSGIGPTTALAI+AVDDNEGLVNAIDNSDIKYLKMPKIGKKT 120
 Query: 121 QQMILDLGKGFVEASGES-TRKVSSEQNSNLEEAMBALALGYKATELKKVKAFFEGT 179
 QQM+LDL+GKFVEA E+ T + + N+ L+EA+EALALGYKA ELKK++AFFEGT
 10 Sbjct: 121 QQMVLDLGKGFVEAPQETGHTKARSNKAGNTQLDEATEALALGYKAKELKKIRAFFEGT 180
 Query: 180 NETVEQYIKSSLKMLMK 196
 +ET EQYIKS+LK+LMK
 15 Sbjct: 181 SETAEQYIKSALKLMLK 197

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1084

A DNA sequence (GBSx1159) was identified in *S. agalactiae* <SEQ ID 3343> which encodes the amino
 20 acid sequence <SEQ ID 3344>. This protein is predicted to be DNA-3-methyladenine glycosidase I (tag).
 Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2812(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 A related GBS nucleic acid sequence <SEQ ID 10263> which encodes amino acid sequence <SEQ ID
 10264> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC76573 GB:AE000432 3-methyl-adenine DNA glycosylase I,
 constitutive [Escherichia coli K12]
 35 Identities = 87/176 (49%), Positives = 122/176 (68%), Gaps = 1/176 (0%)
 Query: 5 MKRCSWVNLNPLYVAYHDKEWGRAVHDDHVLFFELLCLETYQSGLSWE'VTLNKRQEFQV 64
 M+RC WV+ D PLY+AYHD EWG D LFE++CLE Q+GLSW TVL KR+ +R
 Sbjct: 1 MERCGWVSQD-PLYIAYHDNEGWVPETDSKKLFEMICLEGQQAGLSWITVLKKRENYRAC 59
 40 Query: 65 FHHYNIKVAAMSDADLEIILQNPRVIRHLKLFSTRQNARSIIILIQKEFGSFDRIYWSF 124
 FH ++ KVAAM + D+E ++Q+ +IRHR K+ + NAR+ + +++ F ++WSF
 Sbjct: 60 FHQFDPVKVAA MQEEDVERLVQDAGIIRHRGKIQAII GNARAYLQMEQNGEPFVDFVWSF 119
 45 Query: 125 VDNKQVQVNSVNNYNDVPASTILSERLSKDLKKRGFKFVGPTCLYSFIQAAGMVNDH 180
 V+++ QV +++P ST+ S+ LSK LKKRGFKFVG T YSF+QA G+VNDH
 Sbjct: 120 VNHQPVQTQATTLSIPTSTASDALSKALKKRGFKFVGTTICYSFPMQACGLVNDH 175

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3345> which encodes the amino acid
 50 sequence <SEQ ID 3346>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4149(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1212-

An alignment of the GAS and GBS proteins is shown below.

Identities = 114/184 (61%), Positives = 135/184 (72%)

```

5  Query: 3  FHMKRCSWVNDNPLYVAYHDKWGRAVHDDHVLFEELLCLETYQSGLSWETVLNKRQEFR 62
    FHMKRCSWV DN LY YHD EWG+ + DD FELLCLE+YQSGLSW TVL KRQ FR
    Sbjct: 2  FHMKRCSWVPKDNQLYCDYHDLWGQPLDDDRDFEELLCLESYQSGLSWLTVLKKRQAFR 61

10  Query: 63  QVFHHYHNIEKVAAMSDADLEITLQNPRVIRHRLKLFSTRQNARSIILIQKEFGSFDRYIW 122
    VFHHY+I VA + ++ L+NP +IRH+LKL +T NA ++ IQKEFGSF Y+W
    Sbjct: 62  TVFHHYDIASVATFTSEEMADALENPSSIIRHKLKLAATVNNAIAVQKIQKEFGSFSTYLW 121

15  Query: 123 SFVDNKKVQNSVNNYNDVPASTTLSERLSKDLKKRGFKFVGPTCLYSFIQAAGMVNDHEN 182
    +FV K N VN N VPA T LS RL+KDLKKRGFKF+GPT +YSF+QA+G+VNDHE
    Sbjct: 122 NFGVGKPINNLVNQENLVPAQTLSIRLAKDLKKRGFKFLGPTTVYSFMQASGLVNDHEE 181

    Query: 183 ICDF 186
    C F
    Sbjct: 182 ACVF 185

```

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1085

A DNA sequence (GBSx1160) was identified in *S.agalactiae* <SEQ ID 3347> which encodes the amino acid sequence <SEQ ID 3348>. This protein is predicted to be competence-damage inducible protein (cinA). Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have an uncleavable N-term signal seq

```

```

30  ----- Final Results -----
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

35 A related GBS nucleic acid sequence <SEQ ID 10261> which encodes amino acid sequence <SEQ ID 10262> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA84071 GB:Z34303 CinA protein [Streptococcus pneumoniae]
Identities = 194/297 (65%), Positives = 236/297 (79%), Gaps = 1/297 (0%)

```

```

40  Query: 1  MVEGSIPLQNLTLGLAVGIVTSKGVQYQYVLEGGPPSELKPMVMEQVVPILSNNGTKLYSRV 60
    +VEG+IPL N TGLAVGG + GV Y+VLEGGPPSELKPMV+ Q++P L G+KLYSRV
    Sbjct: 121 IVEGAIPLPNETGLAVGKLEVDGVTVVLEGGPPSELKPMVLNQLLPKLMT-GSKLYSRV 179

45  Query: 61  LRFFGIGESQLVTILEDIKQNTDPTIAPYAKVGEVTLRLSTKAENQDEADFKLDSLEKE 120
    LRFFGIGESQLVTIL D+I NQ DPT+APYAK GEVTLRLSTKA +Q+EA+ LD LE +
    Sbjct: 180 LRFFGIGESQLVTILADLIDNQIDPTIAPYAKTGEVTLRLSTKASSQEEANQALDILENQ 239

    Query: 121 ILALKTLNDRKLDLILGYGDNNSMARTVLELLKVNKTITAAESLTAGLFQSQLAEFSG 180
    IL +T + L+D YGYG+ S+A V+E LK Q KTI AAESLTAGLFQ+ +A FSG
    Sbjct: 240 ILDCQTFEGISLRDFCYGYGEETSLASIVVEELKRGKTIAAESLTAGLFQATVANFSG 299

50  Query: 181 ASQVFNGGFTTYSMEAKSQLLGIPKKLQYGVVSHFTAEMAQQARQLLKADFGIGLTG 240
    S +F GGF TYS+E KS++L IP K L+E+GVVS FTA+ MA+QAR ++DFGI LTG
    Sbjct: 300 VSSIFEGGFVTYSLEEKSRMLDIPAKNLEHGVVSEFTAQKMAEQARSKTQSDFGISLTG 359

55  Query: 241 VAGPDELEGYPAGTVFIGIATPEGVSSIKVSIGGKSRSDVRHISTLHAFDLVRRALL 297
    VAGPD LEG+P GTVFIG+A +G IKV+IGG+SR+DVRHI+ +HAF+LVR+ALL
    Sbjct: 360 VAGPDSLEGHEPVGTVFIGLAQDQTEVIKVNIGGRSRADVRHIAVMHAFNLVRKALL 416

```

-1213-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3349> which encodes the amino acid sequence <SEQ ID 3350>. Analysis of this protein sequence reveals the following:

Possible site: 22

5 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.91 Transmembrane 134 - 150 (134 - 150)

10 ----- Final Results -----
 bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

15 >GP:CAA84071 GB:Z34303 ClnA protein [Streptococcus pneumoniae]
 Identities = 286/417 (68%), Positives = 336/417 (79%), Gaps = 1/417 (0%)

Query: 1 MKAEIIAVGTEILTGQIVNTNAQFLSEKMAELGIDVYFQTAVGDNEERLLSVITTSQRS 60
 20 MKAE+IAGVTEILTGQIVNTNAQFLSEK+AE+G+DVYFQTAVGDNE RLLS++ ASQRS
 Sbjct: 1 MKAEIIAVGTEILTGQIVNTNAQFLSEKLAIEIGVDVYFQTAVGDNEVRLLSLEIASQRS 60

Query: 61 NLVILCGGLGPTKDDLTQTLAKYLRKDLVYDEQACQKLDFFAKRKPSRTPNNERQAA 120
 +LVIL GGLG T+DDLTQTLAK+L K LV+D QA +KLD FFA R +RTPNNERQAA
 25 Sbjct: 61 SLVILTGGLGATEDDDLTQTLAKFLGKALVFDPOAQEKLDIFFALRPDYARTPNNERQAA 120

Query: 121 VIEGSIPLPNKTGLAVGGFITVDGISYVVLPGPPSELKPMVNEELVPLLSKQYSTLYSKV 180
 ++EG+IPLPN+TGLAVGG + VDG++YVVLPGPPSELKPMV +L+P L S LYS+V
 30 Sbjct: 121 IVEGAIPLPNETGLAVGGKLEVDGVTYVVLPGPPSELKPMVLNQLLPKMTG-SKLYSRV 179

Query: 181 LRFFGIGESQLVTIVLSDFIENQTDPTIAPYAKTGEVTLRLSTKTENQALADKKLGQLEAQ 240
 LRFFGIGESQLVT+L+D I+NQ DPT+APYAKTGEVTLRLSTK +Q A++ L LE Q
 35 Sbjct: 180 LRFFGIGESQLVTILADLIDNQIDPTLAPYAKTGEVTLRLSTKASSQEANQALDILENQ 239

Query: 241 LLSRKTLEQPLADVIFYGYGEDNSLARETFELLVKYDKTITAAESLTAGLFQSTLASFPG 300
 +L +T EG L D YGYGE+ SLA E L + KTI AAESLTAGLPQ+T+A+F G
 40 Sbjct: 240 ILDCQTFEGISLRDFCYGYGEETSLASIVVEELKRQKTIAAESLTAGLFQATVANFSG 299

Query: 301 ASQVFNGGFVTYSMEEKAKMLGLPLEELKSHGVVSAYTAEGMAEQARLLTGADIGVSLTG 360
 S +F GGFVTYS+EEK++ML +P + L+ HGVVS +TA+ MAEQAR T +D G+SLTG
 45 Sbjct: 300 VSSIFEGGFVTYSLEEKSRMLDIPAKNLEEHGVVSEFTAQMAEQARSKTQSDFGISLTG 359

Query: 361 VAGPDMLEEQPAGTVFIGLATQNKVESIKVLISGRSRLDVRYIATLHAFNMVRKTL 417
 VAGPD LE P GTVFIGLA E IKV I GRSR DVR+IA +HAFN+VRK LL
 50 Sbjct: 360 VAGPDSLEGHFVGTVFIGLAQDQGTETVIVKVNIGGRSRADVRHIAVMHAFNLVRKALL 416

An alignment of the GAS and GBS proteins is shown below.

Identities = 201/299 (67%), Positives = 242/299 (80%)

50 Query: 1 MVEGSIPLQNIATGLAVGGIVTSKGVQYVVLPGPPSELKPMVMEQVVPILSNNGTKLYSRV 60
 ++EGSIPL N TGLAVGG +T G+ Y+VLPGPPSELKPMV E++VP+LS + LYS+V
 Sbjct: 121 VIEGSIPLPNKTGLAVGGFITVDGISYVVLPGPPSELKPMVNEELVPLLSKQYSTLYSKV 180

Query: 61 LRFFGIGESQLVTILEDIKNQTDPTIAPYAKVGEVTLRLSTKAENQDEADFKLDSLEKE 120
 LRFFGIGESQLVT+L D I+NQTDPTIAPYAK GEVTLRLSTK ENQ AD KL LE +
 55 Sbjct: 181 LRFFGIGESQLVTIVLSDFIENQTDPTIAPYAKTGEVTLRLSTKTENQALADKKLGQLEAQ 240

Query: 121 ILALKTLDNRKLDLLYGYGDNNSMARTVLELLKVQNKTTITAAESLTAGLFQSQLEAFSG 180
 +L+ KTL+ + L D+ YGYG++NS+AR ELL +KTITAAESLTAGLFQSLA F G
 60 Sbjct: 241 LLSRKTLEQPLADVIFYGYGEDNSLARETFELLVKYDKTITAAESLTAGLFQSTLASFPG 300

Query: 181 ASQVFNGGFVTYSMEAKSLLGIPKKKLQEYGVVSHFTAEMAQQARQLLKADFGIGLTG 240
 ASQVFNGGF TYSME K+++LG+P ++L+ +GVVS +TAE MA+QAR L AD G+ LTG
 Sbjct: 301 ASQVFNGGFVTYSMEEKAKMLGLPLEELKSHGVVSAYTAEGMAEQARLLTGADIGVSLTG 360

-1214-

Query: 241 VAGPDELEGYPAGTVFIGIATPEGVSSIKVSIGGKSRSDVRHISTLHAFDLVRRALLKI 299
 VAGPD LE PACTVFIG+AT V SIKV I G+SR DVR+I+TLHAF++VR+ LLK+
 Sbjct: 361 VAGPDMLEBQAGTVFIGLATQNKVESIKVLISGRSRLDVRYIATLHAFNMVRKTLKL 419

- 5 SEQ ID 3348 (GBS646) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 2-4; MW 61.6kDa), in Figure 134 (lane 3; MW 57.5kDa + lanes 2 & 4; MW 27kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 131 (lane 5-7; MW 36.6kDa) and in Figure 178 (lane 5; MW 37kDa).

GBS646-His was purified as shown in Figure 229, lane 5.

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1086

A DNA sequence (GBSx1161) was identified in *S.galactiae* <SEQ ID 3351> which encodes the amino acid sequence <SEQ ID 3352>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 59
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.37 Transmembrane 148 - 164 (148 - 164)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3353> which encodes the amino acid
 25 sequence <SEQ ID 3354>. Analysis of this protein sequence reveals the following:

- Possible site: 59
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.37 Transmembrane 148 - 164 (148 - 164)
 30 ----- Final Results -----
 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 35 The protein has homology with the following sequences in the databases:

- >GP:AAD04860 GB:AF069745 RecA protein [Streptococcus parasanguinis]
 Identities = 333/381 (87%), Positives = 356/381 (93%), Gaps = 3/381 (0%)
 40 Query: 1 LAKKLKKNEEITKKFGDERKALDDALKNIEKDFGKGAVMRLGERAEQKVQVMSSGSLAL 60
 +AKK KK ++ITKKFGDER KAL+DALK IEKDFGKG++MRLGERAEQKVQVMSSGSLAL
 Sbjct: 1 MAKKQKLLDDITKKFGDEREKALNDALKLIEKDFGKGSIMRLGERAEQKVQVMSSGSLAL 60
 Query: 61 DIALGAGGYPKGRIIETIYGPSSGKTTVALHAVAQAQKEGGIAAFIDAHALDPAYAAAL 120
 DIALGAGGYPKGRIIETIYGPSSGKTTVALHAVAQAQKEGGIAAFIDAHALDP+YAAAL
 45 Sbjct: 61 DIALGAGGYPKGRIIETIYGPSSGKTTVALHAVAQAQKEGGIAAFIDAHALDPSYAAAL 120
 Query: 121 GVNIDELLSSQPDSEQGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGIDGDSHVGLQ 180
 GVNIDELLSSQPDSEQGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGIDGDSHVGLQ
 Sbjct: 121 GVNIDELLSSQPDSEQGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGIDGDSHVGLQ 180
 50 Query: 181 ARMMSQAMRKLASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYASVRLDVRG 240
 ARMMSQAMRKL ASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYASVRLDVRG
 Sbjct: 181 ARMMSQAMRKLASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYASVRLDVRG 240
 55 Query: 241 TQIKGTGDQKQDSSIGKETKI KVVKNKVAPPFKVAEVEIMYGEGISRTGELVKIASDLDI 300
 TQIKGTGDQKQD+++GKETKIKVVKNKVAPPFK A VEIMYGEGISRTGELVKIA+DLDI

-1215-

Sbjct: 241 NTQIKGTGDKQDTNVGKETKIKVVKNVAPPFKVAMVEIMYGEISRTGELVKIATDLDI 300

Query: 301 IQKAGAWFSYNGEKIGQSENAKRYLADHPELFDLIDKVRVKFGLLESESEESAMAVAS 360
 IQKAGAW+SYNGEKIGQSENAK++LADHPE+FDLID KVRV FGL+E+ E ++

5 Sbjct: 301 IQKAGAWFSYNGEKIGQSENAKFLADHPEIFDLIDHKVRVHFGLEIKDEAVKSLDKTE 360

Query: 361 EE---TDDLALDLNGLIEIED 378
 E +++ LDDL+ IEIED

10 Sbjct: 361 EAPVVEEVTLDLDAIEIED 381

An alignment of the GAS and GBS proteins is shown below.

Identities = 339/379 (89%), Positives = 356/379 (93%), Gaps = 1/379 (0%)

15 Query: 1 MAKKTAKABEITKKFGDERRKALDDALKNIEKDFGKGVMLGERAEQKVQVMSSGSLAL 60
 +AKK KK EEITKKFGDERRKALDDALKNIEKDFGKGVMLGERAEQKVQVMSSGSLAL
 Sbjct: 1 LAKKLKKNBEITKKFGDERRKALDDALKNIEKDFGKGVMLGERAEQKVQVMSSGSLAL 60

Query: 61 DIALGAGGYPKGRIVEIYGPSSGKTTVALHAVAQAQKEGGIAAFIDAHALDPAYAAAL 120
 DIALGAGGYPKGRI+EIYGPSSGKTTVALHAVAQAQKEGGIAAFIDAHALDPAYAAAL

20 Sbjct: 61 DIALGAGGYPKGRIIEIYGPSSGKTTVALHAVAQAQKEGGIAAFIDAHALDPAYAAAL 120

Query: 121 GVNIDELLSSQPDSEGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGIDSHVGLQ 180
 GVNIDELLSSQPDSEGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGIDSHVGLQ

25 Sbjct: 121 GVNIDELLSSQPDSEGLEIAGKLIDSGAVDLVVVDSVAALVPRAEIDGIDSHVGLQ 180

Query: 181 ARMSQAMRKLASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYSSVRLDVRG 240
 ARMSQAMRKLASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFY+SVRLDVRG

30 Sbjct: 181 ARMSQAMRKLASINKTKTIAIFINQLREKVGVMFGNPETTPGGRALKFYASVRLDVRG 240

Query: 241 NTQIKGTGEHKDHNVGKETKIKVVKNVAPPFRAEVEIMYGEISRTGELIKIASDLDI 300
 TQIKGTG+ KD ++GKETKIKVVKNVAPPF+ A VEIMYGEISRTGEL+KIASDLDI

35 Sbjct: 241 TTQIKGTGDKQDSSIGKETKIKVVKNVAPPFKVAEVEIMYGEISRTGELVKIASDLDI 300

Query: 301 IQKAGAWFSYNGEKIGQSENAKRYLADNPAIFDLIDHKVRVHFGMTEDDSPVQSELVEE 360
 IQKAGAW+SYNGEKIGQSENAK+YLAD+P +FDLID KVRV FG+ E +S +S +

40 Sbjct: 301 IQKAGAWFSYNGEKIGQSENAKRYLADHPELFDLIDKVRVKFGLLE-ESESEESAMAVA 359

Query: 361 KNEADDLVLDLDAIEIEE 379
 E DDL LDDLN IEIE+

Sbjct: 360 SEETDDLALDLNGLIEIED 378

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1087

45 A DNA sequence (GBSx1162) was identified in *Sagalactiae* <SEQ ID 3355> which encodes the amino acid sequence <SEQ ID 3356>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2344(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 10259> which encodes amino acid sequence <SEQ ID 10260> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG37358 GB:AF028804 NrpR [Lactococcus lactis subsp. cremoris]
 Identities = 69/132 (52%), Positives = 102/132 (77%)

-1216-

5 Query: 5 MIKIYTISSCTSCCKAKTWLNAHQLPYKEQNLGKESLTRDEILEILTKTESGIESIVSSK 64
 MI IYT SCTSCCKAKTWL+ H +P+ E+NL + L+ EI +IL K + G+E ++SS+
 Sbjct: 1 MITIYTAPSTSCCKAKTWLSYHHIPFNERNLIADPLSTTEISQILQKCDGVEGLISSR 60

Query: 65 NRYAKALNCNIEELSVNEVIDLIQENPRILKSPILIDDKRLQVGYKEDDIRAFLPRSRN 124
 NR+ K L + E++S+++ I +I ENP+I++ PI++D+KRL VGY E++IRAFLPR++R
 Sbjct: 61 NRVFKTLGVDFEDISLSQAIISENPMRRPIIMDEKRLHVGYNNEEIRAFLPRTVRV 120

10 Query: 125 VENAEARLRAAL 136
 +EN ARLR+A+
 Sbjct: 121 LENGGARLRSAL 132

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3357> which encodes the amino acid sequence <SEQ ID 3358>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2569(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 117/132 (88%), Positives = 128/132 (96%)

Query: 5 MIKIYTISSCTSCCKAKTWLNAHQLPYKEQNLGKESLTRDEILEILTKTESGIESIVSSK 64
 MIKIYTISSCTSCCKAKTWLNAH+L YKEQNLGKE LT++EIL IL+KTE+G+ESIVSSK
 Sbjct: 1 MIKIYTISSCTSCCKAKTWLNAHKLAYKEQNLGKEPLTKEEILAILSKTENGVESIVSSK 60

30 Query: 65 NRYAKALNCNIEELSVNEVIDLIQENPRILKSPILIDDKRLQVGYKEDDIRAFLPRSRN 124
 NRYAKAL+C+IEELSV+EVIDLIQ+NPRIKSPILIDDKRLQVGYKEDDIRAFLPRSRN
 Sbjct: 61 NRYAKALDCDIEELSVSEVIDLIQDNPRILKSPILIDDKRLQVGYKEDDIRAFLPRSRN 120

35 Query: 125 VENAEARLRAAL 136
 +EN EARLRAAL
 Sbjct: 121 IENTEARLRAAL 132

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1088

A DNA sequence (GBSx1163) was identified in *S.agalactiae* <SEQ ID 3359> which encodes the amino acid sequence <SEQ ID 3360>. Analysis of this protein sequence reveals the following:

45 Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3097(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:BA04987 GB:AP001511 unknown [Bacillus halodurans]
 Identities = 49/82 (59%), Positives = 64/82 (77%), Gaps = 1/82 (1%)

Query: 1 MGFTDETFRRLDDSN-KVEISETLTAVYRSLEEKGYNPINQIVGYVLSGDPAYVPRYND 59
 M D T++F +++ V++ E I +VY +LEEKGYNPINQIVGY+LSGDPAY+PR+ D
 Sbjct: 1 MSSMDNTMKFNVNNEFPVSDVQVLMVSYEALAEKGYNPINQIVGYLLSGDPAYIPRHKD 60

-1217-

Query: 60 ARNQIRKYERDEIVEELVRYYL 81
 AR IRK ERDE++EELV+ YL
 Sbjct: 61 ARTLIRKLERDELIEELVKSYL 82

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3361> which encodes the amino acid sequence <SEQ ID 3362>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3097(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 80/88 (90%), Positives = 85/88 (95%)

- Query: 1 MGFTDETFRRLDDSNKVEISETLTAVYRSLEEKGYNPINQIVGYVLSGDPAYVPRYNDA 60
 MGFTDETFRF+LDD +K +ISETLTAVY SL+EKGYNPINQIVGYVLSGDPAYVPRYNDA
 20 Sbjct: 1 MGFTDETFRFLDDGDKRQISETLTAVYHSLDEKGYNPINQIVGYVLSGDPAYVPRYNDA 60
 Query: 61 RNQIRKYERDEIVEELVRYYLQNGIDL 88
 RNQIRKYERDEIVEELVRYYLQNGID+
 25 Sbjct: 61 RNQIRKYERDEIVEELVRYYLQNGIDV 88

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1089

- A DNA sequence (GBSx1164) was identified in *S.agalactiae* <SEQ ID 3363> which encodes the amino acid sequence <SEQ ID 3364>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

- 35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1575(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 10257> which encodes amino acid sequence <SEQ ID 10258> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

- >GP:CAB14698 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 82/138 (59%), Positives = 109/138 (78%), Gaps = 1/138 (0%)
 45 Query: 1 MRIMGLDVGSKTVGVAISDPIGFTAQGLEIIKIDEESGNFGFDRLAELVKEYKVDKFFVVG 60
 MRI+GLD+G+KT+GVA+SD +G+TAQG+E IKI+E G++G RL+EL+K+Y +DK V+G
 Sbjct: 1 MRILGLDLGTKTLGVALSDEMGWTAQGIETIKINEAEGDYGLSRSLSELIKDYTIDKIVLG 60
 Query: 61 LPKNNMNTSGPRVEASQAYGDKITELFNLFPVEYQDERLTTVQAERMLVEQADISRGKRRK 120
 PKNMN T GPR EASQ + + +N+PV DERLTT+ AE+ML+ AD+SR KRKK
 50 Sbjct: 61 FPKNMNGTVGPRGEASQTFKAVLETTYNVVVLWDERLTTMAAEKMLI-AADVSRQKRRK 119
 Query: 121 VIDKLAAQLILQNYLDRM 138
 VIDK+AA +ILQ YLD +
 55 Sbjct: 120 VIDKMAAVMILQGYLDSL 137

-1218-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3365> which encodes the amino acid sequence <SEQ ID 3366>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1575 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 114/139 (82%), Positives = 126/139 (90%)

Query: 1 MRIMGLDVGSKTVGVAISDPLGFTAQGLEIIKIDEE SGNFGFDRIARLVKEYKVDKFFVVG 60

MRIMGLDVGSKTVGVAISDPLGFTAQGLEIIKIDEE FGF RL ELVK+Y+V++FV+G

Sbjct: 1 MRIMGLDVGSKTVGVAISDPLGFTAQGLEIIKIDEEKAEFGFTRLEELVKQYQVEQFVIG 60

Query: 61 LPKMNNTSGPRVEASQAYGDKITELFNLPEYQDERLITTVQAERMLVEQADISRGKRKK 120

LPKMNNT+GPRV+AS YG+ I LF LPV YQDERLITTV+A+RML+EQADISRGKRKK

Sbjct: 61 LPKMNNTNGPRVDASITYGNHIEHLFGLPVHYQDERLITVEAKRMLIEQADISRGKRKK 120

Query: 121 VIDKLAAQLILQNYLDRMF 139

VIDKLAAQLILQNYL+R F

Sbjct: 121 VIDKLAAQLILQNYLNRNF 139

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1090

A DNA sequence (GBSx1165) was identified in *S.agalactiae* <SEQ ID 3367> which encodes the amino acid sequence <SEQ ID 3368>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2631 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14697 GB:Z99118 yrzB [Bacillus subtilis]

Identities = 50/94 (53%), Positives = 65/94 (68%), Gaps = 5/94 (5%)

Query: 12 EHQHEVITLVDENGNETLFEILLTIDGREEFGKNYVLLVPAGAEDEQGEIEIQAYSFTE 71

EH + IT+VD+ GNE L E+L T + EEFGK+YVL P +++DE E+EI A SFT

Sbjct: 2 EHGEKNITIVDDQGNEQLCEVLFTFEN-EEFGKSYVLYPIESKDDE--EVEILASSFTP 58

Query: 72 NADGTEGDLQPIPEDSDAEWDMIEEVFNFLDEE 105

N DG G+L PI ++D EWD MIEE N+FL +E

Sbjct: 59 NEDGENGELFPI--ETDEEWD MIEETLNTFLADE 90

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3369> which encodes the amino acid sequence <SEQ ID 3370>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3170 (Affirmative) < succ>

-1219-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 90/98 (91%), Positives = 94/98 (95%)

Query: 7 HDHNHEHQHEVITLVDENGNETLFEILLTIDGREEFGKNYVLLVPAGAEDEQGEIEIQA 66
 H+H ++HQHEVITLVDE GNETLFEILLTIDGREEFGKNYVLLVPAG+EEDE GEIEIQA
 10 Sbjct: 3 HNHENDHQHEVITLVDEQGNETLFEILLTIDGREEFGKNYVLLVPAGSEEDSGEIEIQA 62

Query: 67 YSFTENADGTEGDLQPIPEDSDAEWDMIEEVFNFLDE 104
 YSFTEN DGTEGDLQPIPEDSDAEWDMIEEVFNFLDE
 10 Sbjct: 63 YSFTENEDGTEGDLQPIPEDSDAEWDMIEEVFNFLDE 100

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1091

A DNA sequence (GBSx1166) was identified in *S.agalactiae* <SEQ ID 3371> which encodes the amino acid sequence <SEQ ID 3372>. Analysis of this protein sequence reveals the following:

20 Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2059(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1092

A DNA sequence (GBSx1167) was identified in *S.agalactiae* <SEQ ID 3373> which encodes the amino acid sequence <SEQ ID 3374>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

35 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

40	INTEGRAL	Likelihood = -9.18	Transmembrane	314 - 330 (308 - 334)
	INTEGRAL	Likelihood = -6.21	Transmembrane	279 - 295 (274 - 300)
	INTEGRAL	Likelihood = -6.10	Transmembrane	136 - 152 (135 - 157)
	INTEGRAL	Likelihood = -5.31	Transmembrane	232 - 248 (226 - 253)
	INTEGRAL	Likelihood = -4.73	Transmembrane	163 - 179 (162 - 180)
	INTEGRAL	Likelihood = -3.13	Transmembrane	95 - 111 (94 - 119)
	INTEGRAL	Likelihood = -3.03	Transmembrane	386 - 402 (386 - 405)
45	INTEGRAL	Likelihood = -2.18	Transmembrane	204 - 220 (204 - 221)
	INTEGRAL	Likelihood = -2.13	Transmembrane	40 - 56 (40 - 57)
	INTEGRAL	Likelihood = -1.70	Transmembrane	186 - 202 (182 - 202)

----- Final Results -----

50 bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1220-

A related GBS nucleic acid sequence <SEQ ID 10255> which encodes amino acid sequence <SEQ ID 10256> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3375> which encodes the amino acid sequence <SEQ ID 3376>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 53
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -7.38    Transmembrane    315 - 331 ( 311 - 333)
      INTEGRAL    Likelihood = -6.48    Transmembrane    40 - 56 ( 37 - 61)
      INTEGRAL    Likelihood = -6.10    Transmembrane    278 - 294 ( 274 - 298)
10   INTEGRAL    Likelihood = -5.57    Transmembrane    392 - 408 ( 387 - 410)
      INTEGRAL    Likelihood = -3.98    Transmembrane    186 - 202 ( 184 - 208)
      INTEGRAL    Likelihood = -3.93    Transmembrane    339 - 355 ( 338 - 356)
      INTEGRAL    Likelihood = -2.97    Transmembrane    235 - 251 ( 228 - 253)
      INTEGRAL    Likelihood = -2.44    Transmembrane    166 - 182 ( 166 - 182)
15   INTEGRAL    Likelihood = -2.23    Transmembrane    106 - 122 ( 106 - 125)
      INTEGRAL    Likelihood = -1.81    Transmembrane    83 - 99 ( 83 - 101)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
20   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9179> which encodes the amino acid sequence <SEQ ID 9180>. Analysis of this protein sequence reveals the following:

```

25   Possible cleavage site: 13
   >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -7.38    Transmembrane    243 - 259 ( 239 - 261)
      INTEGRAL    Likelihood = -6.10    Transmembrane    206 - 222 ( 202 - 226)
      INTEGRAL    Likelihood = -5.57    Transmembrane    320 - 336 ( 315 - 338)
30   INTEGRAL    Likelihood = -3.98    Transmembrane    114 - 130 ( 112 - 136)
      INTEGRAL    Likelihood = -3.93    Transmembrane    267 - 283 ( 266 - 284)
      INTEGRAL    Likelihood = -2.97    Transmembrane    163 - 179 ( 156 - 181)
      INTEGRAL    Likelihood = -2.44    Transmembrane    94 - 110 ( 94 - 110)
      INTEGRAL    Likelihood = -2.23    Transmembrane    34 - 50 ( 34 - 53)
35   ----- Final Results -----
      bacterial membrane --- Certainty=0.395(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 200/480 (41%), Positives = 310/480 (63%), Gaps = 1/480 (0%)

Query: 40  ILLYSVLSTLLAIANPLLTYPFANGLQTQNLTYTGLMMTKGQIPYSDVFATGGFLYYVTIAL 99
45   +L +S++ + L IA P LT  ANGLQ+ONLY G+M+TKGQ+PYS F TGG Y+V IAL
Sbjct: 40  LLFFSIIISSLTIAPVPLTDAANGLQSONLYIGMMLTKGQLPYSAFTTGGLFYFVIIAL 99

Query: 100 SYLLGSSIWLLIVQFIAXYVSGIYFYKLVYYVAQSEIVSIGMTLIFYIMNIVLGFGGMYP 159
50   SY LGS++WL+ VQ  +Y+SG+Y YKL+ Y+  + V++ ++ +Y++++ LGFGG+YP
Sbjct: 100 SYYLGSTLWLWVFVQVFCFYLSGLYLYKLINYMTEGFQKVALTFSISYLLSVSLGFGGLYP 159

Query: 160 IQWALPFMFLISLWFLIKFCVDNIVDEAFIFYGILAAFSLFIDPQTLIFWLCSEFVLLTATN 219
      Q A+PF+LIS WFL K+  + DEAFI +G + A ++ IDP TLIFW  + V + + N
Sbjct: 160 TQLAMPFILISANFLTKYFACLVKDEAFILFGFVGALAMLIDPSTLIFWFSFACVTVFSYN 219
55

Query: 220 IKQKQSLRGFYQFLCVVFGMILIAITVGYFMFNLQIISSYIDKAIYPPTYFARTNHSFL 279
      I QK  RGFYQ L  +FGMIL+ YT GYF+ NLQ+++ Y+ + + YPFT+F  N S L
Sbjct: 220 ISQKHLARGFYQLLASIFGMILVFTAGYFILNLQVLNPLYLSQTMIPYPTFFKSGNLSLL 279

Query: 280 LSLAIQIVVLLGSGCLFGLWDFIQNRKKASYQIGLNFACIFIIYAIMAIFSRDFNLYHF 339
      LAIQ+  LG G L G+ + I+ K S ++  + + ++AIFS+D+ YH
Sbjct: 280 FGLAIQLFFALGLGLLTGMENVIRRFKNNSDRVVKWLFVMVILESILVAIFSQDYRPHYH 339
60

```

-1221-

Query: 340 LPALPFGLLLTNSKITILYQKVIDRRSHRRQY-FSGKSLIVDLFVKKTYLPLLLVSLSI 398
 LP LPFGL+LT+ + Y + + SHRR++ +G ++ +++K+ +YLP+L+V +
 5 Sbjct: 340 LPLLPFGLLILTAIPVGYYQYIGLQSSHRRRHGKNGVGRVMMIYLRHFFYLPIILVGTIL 399

Query: 399 GLIVYNTYQNVTLSEKERRDISHYLTTKIDRDGKIYVWDKVASIYSQTRLKSASQFVLPHI 458
 Y ++ L++ER I+ YL K+++ IYVD + IY ++ KS SQF P I
 Sbjct: 400 ICSTYCFISSIPLNQERDHIASYLEQKLNKTQSIYVWDDTSKIYLDKAKSVSQFSSPDI 459

Query: 459 NTAQKNNEKILKDELLQHGAKYFILNKNEKLPNELKSDIKKHQEVPLSNITHFVLYRFX 518
 NT ++++ KIL+DELL++ A Y ++N+ + LP ++ + +Y+ F++Y+ K
 10 Sbjct: 460 NTQKESHRKILEDELLENKAAIYIVNRYKNLPKIIQKVLSTNYKVDKQITTKSFIVYQKK 519

15 A related GBS gene <SEQ ID 8727> and protein <SEQ ID 8728> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0
 SRCFLG: 0
 McG: Length of UR: 34
 Peak Value of UR: 2.23
 20 Net Charge of CR: 0
 McG: Discrim Score: 7.72
 GvH: Signal Score (-7.5): -2.21
 Possible site: 60
 >>> Seems to have a cleavable N-term signal seq.
 25 Amino Acid Composition: calculated from 61
 ALOM program count: 5 value: -9.18 threshold: 0.0
 INTEGRAL Likelihood = -9.18 Transmembrane 174 - 190 (168 - 194)
 INTEGRAL Likelihood = -6.21 Transmembrane 139 - 155 (134 - 160)
 INTEGRAL Likelihood = -5.31 Transmembrane 92 - 108 (86 - 113)
 30 INTEGRAL Likelihood = -3.03 Transmembrane 246 - 262 (246 - 265)
 INTEGRAL Likelihood = -2.18 Transmembrane 64 - 80 (64 - 81)
 PERIPHERAL Likelihood = 3.29 194
 modified ALOM score: 2.34
 icml HYPID: 7 CFP: 0.467
 35 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF02392(331 - 978 of 1764)
 45 EGAD|43696|MJ1079(2 - 379 of 397) conserved hypothetical protein {Methanococcus jannaschii}
 OMNI|MJ1079 conserved hypothetical protein GP|1591727|gb|AAB99076.1||U67550 conserved
 hypothetical protein {Methanococcus jannaschii} PIR|F64434|F64434 hypothetical protein
 MJ1079 - Methanococcus jannaschii
 %Match = 3.1
 50 %Identity = 25.6 %Similarity = 50.7
 Matches = 57 Mismatches = 100 Conservative Sub.s = 56

174 204 234 264 294 324 354
 *ILLANI*LSVHPTSFFTXXXN*LXXSSIWLLIVQFIAYVSGIYFKLVYYVAQSEIVSIGMTLIFYIMNIVLG-----
 55 : |:: |:: |:: |
 MLNLLYLILGIICGTITGL
 10

426 447 477 507 537 567 597
 60 FGGMYPIQW-ALPFMLISLWFL---IKFCVDNIVDEAFIFYGILAAFSLFIDPQTLIFWLCFSVLLTATNIKQKQSLRGF
 | |::| || |::: | | | : :: || : |::| | : : | | : |::|
 FPGIHPPNIVALSFLILPYFGLDNYIFFLIGLVITHYFINF-IPSAFLGVDPDETAVALPMHKLTLNNGNGYEAVIAGF
 30 40 50 60 70 80 90

627 657 687 717 747 774
 65

-1222-

```

5      YQFLCVVFGMILIAITVGYFMFNLIQIISSYIDKAIFFYPPTYFARTNHSFLLSLAI-QIVVLLGSGC-----
      :| ||| : : : :|:: : | || || | : : : :|:: :|
      GSYLGVVFSILISLFLMSILHFDVRAFYCSI--KIFIPFILIAFILIYQIFTAKSVWEVLVIFLSGIFGIAVLYCSEAFNI
      110      120      130      140      150      160      170

10     798      828      846      876
      -----LFGLWDFIQNRKKASYQ-----IGLNFIAICIFI
      :||: :| | | : : :|:: :|
      TLTAIFTGMFGIPLLINNLTQYIKISQMMAFDPDFELKFLKSSFFA~~~~TIAIILLNLNLSKYILLFIRKVNFKFLSLFFI
      190      200      210      220      320      330

15     906      948      978      1008      1038      1068      1098
      IYAIMAIFSRDFN---LYH---FLPALPFGLLLTSNKITILYQKVIDRRSHRRQYFSGKSLIVDLFVKKTYLPLLLVSL
      |: : : :| :|| :| :| | : : :
      IFCSLVVIIGSYNTYLIYHIIVLTAIYIGLLAVKSNNTNLSNMMNVLIPTTILYFLRG
      350      360      370      380      390

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 1093

A DNA sequence (GBSx1168) was identified in *S.agalactiae* <SEQ ID 3377> which encodes the amino acid sequence <SEQ ID 3378>. This protein is predicted to be anaerobic ribonucleotide reductase (nrdD). Analysis of this protein sequence reveals the following:

```

25     Possible site: 52
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3722(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10253> which encodes amino acid sequence <SEQ ID 10254> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

35     >GP:AAD00215 GB:U73336 anaerobic ribonucleotide reductase
      [Lactococcus lactis subsp. cremoris]
      Identities = 539/725 (74%), Positives = 616/725 (84%), Gaps = 7/725 (0%)

40     Query: 10 MTESDIKVIKRDGRVLSFDKYKIYTALLKASNKVIKMSPLVEAKLEMIADHVIAEIYNRF 69
      +T +I VIKRDGR V F+ KI+ AL KA+ KV V L + D V++EI++RF
      Sbjct: 10 VTLEEINVIKRDGRSVKFNSEKIFDALTKAAKKVELTDKSV---LSELTDRVSEIFSRF 66

      Query: 70 KDNIKIYEIQNIVEHKLLEANEYATAQEYINRYRTQRDFERSQATDINFSGKLINKDQTV 129
      +N+KIYEIQ+IVE +LLE+ E A+A+EYI+YR RD R++ATDINF+I KLIN+DQTV
45     Sbjct: 67 SENVKIYEIQSIVEQEELLESETALAEYISYRANRDLARTKATDINFTEKLINRDQTV 126

      Query: 130 VVENANKDSDFVNTQRDLTAGIVGKSIGLKMLPSHVANAHQKGDIIHYHDLDYSPYTPMTN 189
      VVENANKDS+VFNTQRDLTAG V K+IGLK+LP HVANAHQKGDIIHYHDLDYSP+T M N
50     Sbjct: 127 VVENANKDSNVFNTQRDLTAGAVSKAIGLKLPPHVANAHQKGDIIHYHDLDYSPFTTMAN 186

      Query: 190 CCLIDFPKGLMANGFKIGNAEVSPKSIQTATQAISQIIANVASSQYGGCTADRIDEFLAP 249
      CCLIDFPK M NGFK+GNA+V+SPKSIQTATQA SIIANVASSQYGGC+ DR DE LAP
      Sbjct: 187 CCLIDFPKNMFENGFKLGNAQVDSPKSIQTATQAASQIIANVASSQYGGCSFDRADEVLAP 246

55     Query: 250 YAQLNYQKHLKDAKEWVIED-KQEDYARAKTQKDIYDAMQSLEYEINTLFTSNGQTPPTS 308
      YA+LNYQKHLKDA++W+ D K+E YAR KT KDIYDAMQSLEYEINTLFTSNGQTPF +
      Sbjct: 247 YAKLNYQKHLKDAQKWKIDGDEKREAYAREKTAKDIYDAMQSLEYEINTLFTSNGQTPFVT 306

      Query: 309 LGFGLGTNWFEREIQKAILKIRIQGLGSEHRTAIFPKLIFTLKKGLNLEEDSPNYDIKQL 368

```

-1223-

+GFGLG +W+ REIQKAILK+RI GLGSEHRTAIFPKLIFTLK+GLNLE +PNYDIK+L
 Sbjet: 307 VGFGGLGDDWYAREIQKAILKVRIGGLGSEHRTAIFPKLIFTLKRLNLEVGTPNYDIKEL 366
 Query: 369 ALECATKRMYPDVLSDYDKIIDLTGSEFKAPMGCRSFLQGWRDANGQDVTSGRMNLGVVTVN 428
 ALEC+TKRMYPD+LSYDKI++LTGSEFKA MGRSFLQGW+DANG DVT+GR NLGVVTVN
 Sbjet: 367 ALECASTKRMYPDILSYDKIVELTGSEFKASMGCRSFLQGWKDANGNDVTAGRNNLGVVTVN 426
 Query: 429 LPRVAMESNGDMDKFWEIFNERMSIARDALVYRVERVKEAIPANAPILYQYGAFGERLGK 488
 LPR+A+E+ G+ +KFWEIFNER+ IA DAL +RVER KEA P NAPIL+ GA G RL
 Sbjet: 427 LPRIALEAAGNKEKFWEIFNERVEIAHDALAFRVERAKEAQPKNAPILFMNGALG-RLDS 485
 Query: 489 YDNVDRLFNHRRATVSLGYIGLYEVASVFGGDWEDNHQAKAFTVDIVRKMQLCADWSD 548
 +VD L+N+ RATVSLGYIGLYEVA+ FYG WE N +AKAFT++IV++M + C DWS
 Sbjet: 486 EGSVDDLYNNERATVSLGYIGLYEVATTFFYGPTWESNPEAKAFTIETVKRMHEDCEDWSK 545
 Query: 549 EYDYHFSVYSTPSESLTDRFCRLDTEKFGIVTDITDKEYYTNSFHYDVRKNPTPFKEKLD 608
 YH+SVYSTPSESLTDRFCR+D EKFG V DITDK+YYTNSFHYDVRKNPTPFKEKLD+F
 Sbjet: 546 ASGYHYSVYSTPSESLTDRFCRMDKEKFGSVADITDKDYTTNSFHYDVRKNPTPFKEKLEF 605
 Query: 609 EKIYPETGASGGFIHYCEYPVLQQNPKALEAVWDYAYDRVGYLGTNTPIDKCYQCQFEGD 668
 EK YP A+GGFIHYCEYPVLQQNPKALEAVWD+AYDR+GYLGTN PID CY C FEGD
 Sbjet: 606 EKDYF-VYANGGFIHYCEYPVLQQNPKALEAVWDFAYDRIGYLGTPIDHCYACGFEGD 664
 Query: 669 FTPTDRGFTCPNCGNSDPKTDVVKRTCGYLGNPQARPMVNGRHKESARVKHMNGS-SI 727
 FTPT+RGF CP CGN DPKT DVVKRTCGYLGNPQARPMV+GRHKES+RVKHMNGS
 Sbjet: 665 FTPTERGFKCPQCGNDPKTCDVVKRTCGYLGNPQARPMVHGRHKES+SRVKHMNGSVGA 724
 Query: 728 KNQCN 732
 N GN
 Sbjet: 725 LNDGN 729

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3379> which encodes the amino acid sequence <SEQ ID 3380>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2975 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 641/731 (87%), Positives = 680/731 (92%)

Query: 1 MMVLERERFMTESDIKVIKRDGRLVSFDKYKIYTALLKASNKVIKMSPLVEAKLEMIADH 60
 M+ LE ++ + DIKVIKRDGRLV+FD KIY+ALLKAS KV +MSPLVEAKLE I+D
 Sbjet: 1 MVSLEEDKVTVQPDIKVIKRDGRLVNFDSTKIYSALLKASMKVTRMSPLVEAKLEAISDR 60
 Query: 61 VIAEIYNRFKDNIKIYEIQNIVEHKLLEANEYAIQAQYINRYRTQDRDFERSQATDINFSIG 120
 +IAEI RF NIKIYEIQNIVEHKLLEANEYAIQAQYINRYRTQDRF RSQATDINFSI
 Sbjet: 61 IIAEIIERFPTNIKIYEIQNIVEHKLLEANEYAIQAQYINRYRTQDRFARSQATDINFSID 120
 Query: 121 KLINKDQTVVNENANKSDVFNTQDRDLTAGIVGKSI GLKMLPSHVANAHQKGDIIHYHDL 180
 KLINKDQTVVNENANKSDVFNTQDRDLTAGIVGKSI GLKMLPSHVANAHQKGDIIHYHDL
 Sbjet: 121 KLINKDQTVVNENANKSDVFNTQDRDLTAGIVGKSI GLKMLPSHVANAHQKGDIIHYHDL 180
 Query: 181 YSPYTPMTNCCLIDFKGMLANGFKIGNAEVESPKSIQTATAQISQIIANVASSQYGGCTA 240
 YSPYTPMTNCCLIDFKGMLANGFKIGNAEVESPKSIQTATAQISQIIANVASSQYGGCTA
 Sbjet: 181 YSPYTPMTNCCLIDFKGMLANGFKIGNAEVESPKSIQTATAQISQIIANVASSQYGGCTA 240
 Query: 241 DRIDEFLAPYAQLNYQKHLKDAKEWVIEDKQEDYARAKTQKDIYDAMQSLEYEINTLFTS 300
 DRIDEFLAPYA+LN++KH+ DAK+W++E K+E YA KTQKDIYDAMQSLEYEINTLFTS
 Sbjet: 241 DRIDEFLAPYAELNFKKHMADAKKWIVETKRESYAFKTKQDIYDAMQSLEYEINTLFTS 300
 Query: 301 NGQTPFTSLGFGLTNWFEREIQKAILKIRIQGLGSEHRTAIFPKLIFTLKKGLNLEEDS 360

-1224-

5 NGQTPFTSLGFLGT+WFEREIQKAIL IRI GLGSEHRTAIFPKLIFT+K+GLNLE DS
 Sbjct: 301 NGQTPFTSLGFLGTSWFEREIQKAILTIRINGLGSEHRTAIFPKLIFTVKRGLNLEPDS 360
 Query: 361 PNYDIKQLALECATKRMYPDVLSYDKIIDLTGSFKAPMGCRSFLQGWDRDANGQDVTSGRM 420
 PNYDIK LALECATKRMYPD+LSYDKIIDLTGSFK+PMGCRSFLQGW+D NGQDVTSGRM
 Sbjct: 361 PNYDIKTLALECATKRMYPDMLSYDKIIDLTGSFKSPMGCRSFLQGWKDENGQDVTSGRM 420
 Query: 421 NLGVVTVNLPRVAMESNGDMDKFWEIFNERMSIARDALVYRVERVKEAIPANAPILYQYG 480
 NLGVVT+NLPR+AMESNGDMDKFWE+FNERM I++DAL+YRVERV EA PANAPILYQYG
 10 Sbjct: 421 NLGVVTVNLPRAMESNGDMDKFWEFNERMLISKDALIYRVERVTEAKPANAPILYQYG 480
 Query: 481 AFGERLKYDNDVDRLFNHRATVSLGYIGLYEVASVFYGGWEDNHQAKAFTVDIVRKMK 540
 AFG+RL K NV+ LF +RRATVSLGYIGLYEVASVFYGG WE N AKAFT+ IV+ MK
 Sbjct: 481 AFGKRLEKTGNVNDLFKNRRATVSLGYIGLYEVASVFYGGWEGNPDAKAFTLSIVKAMK 540
 15 Query: 541 QLCADWSDEYDYHFSVYSTPSESITDRFCRLDTEKFGIVTDITDKEYYTNSFHYDVRKNP 600
 Q C DWSDEY YHFSVYSTPSESITDRFCRLDTEKFGIVTDITDKEYYTNSFHYDVRK+P
 Sbjct: 541 QACEDWSDEYGYHFSVYSTPSESITDRFCRLDTEKFGIVTDITDKEYYTNSFHYDVRKSP 600
 Query: 601 TPFEKLDPEKIYPETGASGGFIHYCEYPVLQONPKALEAVWDYAYDRVGYLGTNTPIDKC 660
 TPFEKLDPEK YPE GASGGFIHYCEYPVLQONPKALEAVWDYAYDRVGYLGTNTPIDKC
 Sbjct: 601 TPFEKLDPEKDYPEAGASGGFIHYCEYPVLQONPKALEAVWDYAYDRVGYLGTNTPIDKC 660
 20 Query: 661 YQCQFEGDFTPTDRGFTCPNCGNSDPKTVDVVKRTCGYLGPNQARPMVNGRHKEISARVK 720
 Y CQFEGDFTPT+RGFTCPNCGN+DPKTVDVVKRTCGYLGPNQARPMVNGRHKEISARVK
 Sbjct: 661 YNCQFEGDFTPTERGFTCPNCGNNDPKTVDVVKRTCGYLGPNQARPMVNGRHKEISARVK 720
 25 Query: 721 HMNGSSIKNQG 731
 HMNGS+IK G
 Sbjct: 721 HMNGSTIKYPG 731
 30

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1094

35 A DNA sequence (GBSx1169) was identified in *S.agalactiae* <SEQ ID 3381> which encodes the amino acid sequence <SEQ ID 3382>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5372(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3383> which encodes the amino acid sequence <SEQ ID 3384>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.6084(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below.

Identities = 28/47 (59%), Positives = 40/47 (84%), Gaps = 1/47 (2%)

Query: 1 MGKYQLDYKGQAQVQKFHEKHSTGENANQKSRLLKDLRKQFLEKAKKK 47

-1225-

MGKYQLDYKG QV++FHEKHS + ++KSR+++L+ +FLEK+KK+
 Sbjct: 1 MGKYQLDYKGMQOVERFHEKHHSK-KKTDKKSrvQELKARFLEKSKKQ 46

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1095

A DNA sequence (GBSx1170) was identified in *S.galactiae* <SEQ ID 3385> which encodes the amino acid sequence <SEQ ID 3386>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0436(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB95794 GB:AL359949 putative oxidoreductase [Streptomyces
 coelicolor A3(2)]
 Identities = 91/299 (30%), Positives = 147/299 (48%), Gaps = 7/299 (2%)

Query: 2 LQLGIVGLGGISQKAYLPYMRQVTGVHWHLFTRQKQILEEV--NMLFGSSTAYDSLDSLA 59
 +++G +GLG I+QK YLP + + G+ HL TR L V + + + LD+L
 Sbjct: 1 MKVGCIGLGDIAQKGYLPVLAALPGIELHLQTRTPATLTRVADKLRIPPAQRHADLDALL 60

Query: 60 EHPLDGVFIHVATSAHFDAKFLKKGIPVFMKPLTEDYTSTKALYDLAKDHKTFLMAG 119
 LD F+H T+AH +I L+ G+P ++DKPL + ++ L LA++ T L G
 Sbjct: 61 AQGLDAAFVHAPTAHPEIVTRLLEAGVPTYVDKPLAYELADSERLVTLEERGTSLAVG 120

Query: 120 FNRRFAPRIMEMKKVEDKNHIRTFFKNAVNAPADFQYKLFDMFIHPLDTALFLTNVVKRG 179
 FNRR AP + + + I KN P D + + D FIH +DT FL V
 Sbjct: 121 FNRRHAPGYAQCAE-HPRELILMQKNRTGLPEDPRTMILDDFIHVVDTLRFLVPGPVDDV 179

Query: 180 YFVTKRDGNKILQVSVTLETDSEIIEASMNQSGSRREIIEIESPEVTYSLDDLSNLSVI 239
 + +G + V + L D MN SGS EI+E+ + + +L+ VI
 Sbjct: 180 TVRARTEGGLHHVVLQLAGDGFTALGVMNRLSGSABEILEVSGQDTKRQVNLNLA--EVI 237

Query: 240 DGFDRRAI-GFGSWASTLEKRGFEPMDAFIQAITTGVPNPISPKSLLSHFICDQINKA 297
 D + + G W +RG E + AF+ A+ +G +S + +L +H +C+++ +A
 Sbjct: 238 DHKGQPTVRRRGDWVPVARQRCIEQAVLAFLDAVRSG-EVLGARDALATHELCERVVRA 295

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3387> which encodes the amino acid sequence <SEQ ID 3388>. Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF96942 GB:AB004430 oxidoreductase, Gfo/Idh/MocA family [Vibrio cholerae]
 Identities = 103/304 (33%), Positives = 158/304 (51%), Gaps = 11/304 (3%)

Query: 4 LNIGIVGLGAISQKAYLPYMRQLSDITWHLSTRNAAVRQQVGLFGHAILYSDEVKELSKT 63
 + I ++GLG I+QKAYLP + Q DI L TRN V + + + +D +++ +
 Sbjct: 1 MKIAMIGLGDIAQKAYLPVLAQWPDIELVLCITRNPKVLGTLATRYRVSATCTDYRDVLQY 60

-1226-

Query: 64 NLDGVFIHAATSAHAELASLFLNQGIPVFMKDPIADNYLMTKNLYDLAKENQTFMLMAGFN 123
 +D V IHAAT H+ LA+ FL+ GIP F+DKP+A + +NLY+LA+++ L GFN
 Sbjct: 61 GVDAMIHAATDVHSTLAAFFLHLGIPTFVDPKPLAASAQECENLYELAEKHHQPLYVGFN 120

5 Query: 124 RRFTPRVKK-LSSLSTK-----RKVAVEKNDLNRPGDMTFKLPDFFIHPLDTALFLTEGT 177
 RR P + LS L+ + R + EK+ PGD+ +FD FIHPLD+ +
 Sbjct: 121 RRHIPLYNQHLSELAQQECCGALRSLRWEKRRHALPGDIRTFVDFDFIHPDLDSVNLRSQC 180

10 Query: 178 LLKGHFQYHLEAGLLSQVMVTLMTESMTTASMNLSQSGSRREVMVQRAEETYLHLENLDE 237
 L H YH+ GLL+++ V T ASMN Q G E + Y ++ +
 Sbjct: 181 LDDLHLTYHMSSEGLLARLDVQWQTGDTLLHASMNRFGITTEHVTASYDNVAYLFDSFTQ 240

15 Query: 238 LSIYKGTETKRVLGFAWDTTLHKRGFETMIDAFLEAISTGVNPVS-PESSLLSHW----I 292
 +++ ++ + W L +GF+ M+ +L+ + G P E +L SH I
 Sbjct: 241 GKMWRDNQESRVALKDWTPMLASKGFDAMVDWLQVAAAGKLPHTIERNLASHQLAEAI 300

20 Query: 293 CQQI 296
 CQQI
 Sbjct: 301 CQQI 304

An alignment of the GAS and GBS proteins is shown below.

Identities = 168/308 (54%), Positives = 223/308 (71%)

25 Query: 1 MLQLGIVGLGGISQKAYLPYMRQVTGVHWHLFTRQKQILEEVNMLFGSSTAYDSLSLAE 60
 ML +GIVGLG ISQKAYLPYMRQ++ + WHL TR + ++V LFG + Y + L++
 Sbjct: 3 MLNIGIVGLGAI SQKAYLPYMRQLSDITWHLSTRNAAVRQQVQQLFGHAILLYSDVKELSK 62

30 Query: 61 HPLDGVFIHVATSAHFDAIKLFLKKGIPVFMKDPLTETYSTKALYDLAKDHKTFLMAGF 120
 LDGVFIH ATSAH ++A LFL +GIPVFMKDGP+ ++Y TK LYDLAK+++TFLMAGF
 Sbjct: 63 TNLDGVFIHAATSAHAELASLFLNQGIPVFMKDPIADNYLMTKNLYDLAKENQTFMLMAGF 122

35 Query: 121 NRRFAPRIMEMKKVEDKNHIRTFFKNVAVNAPADFQYKLFDMFIHPLDTALFLTNVVKRGY 180
 NRRF PR+ ++ + K + KN +N P D +KLFD FIHPLDTALFLT + +G+
 Sbjct: 123 NRRFTPRVKLSSLSTKRKVAVEKNDLNRPGDMTFKLPDFFIHPLDTALFLTEGTLLKGH 182

40 Query: 181 FVTKRDNKILQVSVTLETDSIIIEASMNLSQSGSRREIIIESPEVTYSLDDLNLNLSVID 240
 F + + QV VTL T+S ASMNLSQSGSRRE++E++ E TY L++L LS+
 Sbjct: 183 FQYHLEAGLLSQVMVTLMTESMTTASMNLSQSGSRREVMVQRAEETYLHLENLDELSIYK 242

45 Query: 241 GFDRAIGFGSWASTLEKRGFEPMIDAFIQAITGVNPISPSSLLSHFICDQINKANAP 300
 G ++R +GF SW +TL KRGFE MIDAF++AI+TGVPN+SP+SSLLSH+IC QI +
 Sbjct: 243 GTEKRVLGFAWDTTLHKRGFETMIDAFLEAISTGVNPVSPSSLLSHWICQIADSQLS 302

Query: 301 FGMLNLKI 308
 +G L +++
 Sbjct: 303 YGELTVEL 310

SEQ ID 3386 (GBS309) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 10; MW 63kDa).

50 GBS309-GST was purified as shown in Figure 212, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1096

55 A DNA sequence (GBSx1171) was identified in *S.galactiae* <SEQ ID 3389> which encodes the amino acid sequence <SEQ ID 3390>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----

-1227-

bacterial cytoplasm --- Certainty=0.2983(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04222 GB:AP001508 unknown conserved protein in others

[Bacillus halodurans]

Identities = 52/129 (40%), Positives = 70/129 (53%), Gaps = 5/129 (3%)

10 Query: 39 FEDWLDHNLNMBELGVGVPDNFVPIQFVSFDNDNNAIGFLNLRRLNDTLEKGGHIGYS 98
 FE L + + GV +P N V + IG +N+R LND L +GGHIGY
 Sbjct: 43 FEHLKTLKDYQHGVLNLPANRVANTTYWLVHEQRLIGAINIRHTLNDWLHHRGGHIGY 102

15 Query: 99 IRPRQRGKGYAKEQLKLGIEQAHLKNINEILVTCCHVDNDASKSVILANGGVLEDCLHQ-- 156
 IRP +RGKGYA LKLG+E+A + ++L+TC +N S I NGGVL+ +
 Sbjct: 103 IRPSERGKGATMLKLGLEKAAALGLEKVLITCDKENLPSARTIQRNGGVLDSEVVDER 162

Query: 157 ---TERYWI 162
 +RYWI

20 Sbjct: 163 GIAIQRYWI 171

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3391> which encodes the amino acid sequence <SEQ ID 3392>. Analysis of this protein sequence reveals the following:

Possible site: 59
 25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2195(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 90/164 (54%), Positives = 115/164 (69%), Gaps = 4/164 (2%)

35 Query: 1 MKLRRPVLEDKEEILAMYKEFQKESSSVDG--GFYEPTMHFEDWLDHNLNMBELGVGVPDN 58
 M+RRP L+DK+ +L+M EF ++ S+ DG F ++E WL+ +L E+G+
 Sbjct: 1 MEIRRTLKDKDAVLSMINEFLEQKSATDGLWHFNVNDFNYETWLEDLSLRQEMGLS--SQ 58

40 Query: 59 FVPYIQFVSFDNDNNAIGFLNLRRLNDTLEKGGHIGYSIRPRQRGKGYAKEQLKLGIE 118
 VP IQ+V+FD + AIGFLNLRRLN+ LLEKGGHIGYS+RP QRGKGYAKE LK +
 Sbjct: 59 GVPATQYVAFDERSQAIGFLNLRRLNERLLEKGGHIGYSVRPSQRGKGYAKEMLKQAVS 118

Query: 119 QAHLKNINEILVTCCHVDNDASKSVILANGGVLEDCLHQTERYWI 162
 A KNI ILVTC N AS++VI+AN G+LED TERYWI

45 Sbjct: 119 YAISKNITTILVTCDETNVASRAVIVANVGILEDSRGGTERYWI 162

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1097

50 A DNA sequence (GBSx1172) was identified in *S.agalactiae* <SEQ ID 3393> which encodes the amino acid sequence <SEQ ID 3394>. This protein is predicted to be anaerobic ribonucleotide reductase activator protein (nrdG). Analysis of this protein sequence reveals the following:

Possible site: 59
 55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4239(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1228-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAD00216 GB:U73336 anaerobic ribonucleotide reductase activator
protein [Lactococcus lactis subsp. cremoris]
Identities = 152/198 (76%), Positives = 176/198 (88%)

10 Query: 8 NTPKPGEWKSEELSHGHIIDYKAFNFVDGEGVRNSLYVAGCMFHC KGCYNTATWSFRAGI 67
N PKPGEW+++ELS +I DYK FNFVDGEGVR SLYV+GCMFHC+GCYN ATWSFR G
Sbjct: 2 NNPKPGEWRADELSQNYIADYKPFNFVDGEGVRCSLYVSGCMFCEGCYNQATWSFRYGR 61

15 Query: 68 PYTKLEEDQIMTDLEQPYVQGLTLLGGEPFLNTGILLPLLQIRRELPEKDIWSWTGYTW 127
PYTKLEED+IM DL +PYVQGLTLLGGEPFLNT L+PLL+RIRREL+KDIWSWTGYTW
Sbjct: 62 PYTKLEEDKIMADLAEPYVQGLTLLGGEPFLNTTFLIPLLKRIRREL+KDIWSWTGYTW 121

20 Query: 128 EEMMLETQDKLEMLSLIDILVDGRFDQSKRNMLQFRGSSNQRIIDVQKSLKEGEVVIWE 187
EEMMLET DKLEML L+D+LVDGRF+ SK+NLMLQFRGSSNQRIIDV KS +G+VVIWE
Sbjct: 122 EEMMLETDDKLEMLDLLDLVDGRFELSKKNMLQFRGSSNQRIIDVPKSRSKGQVVIWE 181

25 Query: 188 GLNDGDNSYEQVKRDDL 205
LNDG+N++EQ+ ++ L+
Sbjct: 182 KLNDGENNFEQIHKEKLI 199

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3395> which encodes the amino acid
sequence <SEQ ID 3396>. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4111(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 167/202 (82%), Positives = 186/202 (91%)

40 Query: 4 EASWNTPKPGEWKSEELSHGHIIDYKAFNFVDGEGVRNSLYVAGCMFHC KGCYNTATWSF 63
E WN PKP EW++EELS G IIDYKAFNFVDGEGVRNSLYV+GC+FHCKGCYN ATWSF
Sbjct: 4 EKCWNPKPKQAEELSQGRIIDYKAFNFVDGEGVRNSLYVSGCLFHCKGCYNAATWSF 63

45 Query: 64 RAGIPYTKLEEDQIMTDLEQPYVQGLTLLGGEPFLNTGILLPLLQIRRELPEKDIWSWT 123
+AG+PYT+ELE+QIMTDL QPYVQGLTLLGGEPFLNTGIL+PL++RIRRELPEKDIWSWT
Sbjct: 64 KAGMPYTQEELEEQIMTDLAQPYVQGLTLLGGEPFLNTGILIPLIKRIRRELPEKDIWSWT 123

50 Query: 124 GYTWEEMMLETQDKLEMLSLIDILVDGRFDQSKRNMLQFRGSSNQRIIDVQKSLKEGEV 183
GYTWEEMMLET DKLEMLSLIDILVDGRFD +K+NLMLQFRGSSNQRIIDVQKSL EV
Sbjct: 124 GYTWEEMMLETPDKLEMLSLIDILVDGRFDITKKNMLQFRGSSNQRIIDVQKSLAAKEV 183

Query: 184 VIWEGLNDGDNSYEQVKRDDL 205
+IW+ LNDGD ++EQ+ R+DLL
Sbjct: 184 IIWDKLNDGDQTFEQISREDLL 205

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1098

A DNA sequence (GBSx1173) was identified in *S.agalactiae* <SEQ ID 3397> which encodes the amino acid sequence <SEQ ID 3398>. Analysis of this protein sequence reveals the following:

Possible site: 27
>>> Seems to have no N-terminal signal sequence

-1229-

INTEGRAL Likelihood = -3.03 Transmembrane 102 - 118 (101 - 119)

----- Final Results -----

5 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAD24446 GB:AF118389 unknown [Streptococcus suis]
Identities = 97/240 (40%), Positives = 151/240 (62%), Gaps = 1/240 (0%)

Query: 2 IKILIPTAKEMKV-CQNIAPKLSAQTKIIDIYFSTLTVDLEDIYRINTSAARCEAQRW 60
+KI+IP AKE+ +N ++ LS ++K ++D S V + Y++N + A EA RW
Sbjct: 1 MKIIPNAKEVNTNLENASFYLLSDRSKPVLDALISQFDVKKMAAFYKLNEAKAELEADRW 60

15 Query: 61 QDFKAKQLTLNPAIKLFNGLMYRNIKRHNLSSEAQFMENSVFITSALYGIIPAMTLISP 120
+ Q PA +L++GLMYR + R + + E ++ + V + +ALYG+I ISP
Sbjct: 61 YRIITGQAKTYPAWQLYDGLMYRMDRRGIDSKENYLRDHVRVATALYGLIHPFEFISP 120

20 Query: 121 HRLDFNTKIKINNNSLKVFWRENYDTFMQSDDIMVSLLSNEFETVFSPEKQKLIHLNFI 180
HRLDF +KI N SLK +WR YD + D+++SL S+EFE VFSP+ +++L+ + F+
Sbjct: 121 HRLDFQGSLSKIGNQSLKQYWRPYDQEVGDDELILSLASSEFEQVFSPIQKRLVKILFM 180

25 Query: 181 EDRDGLKTHSTISKARGKCLTAMMENNCQTLEHLKQLRFDGFCYDNELSDSKQLTFVK 240
E++ GQLK HSTISKK RG+ L+ + +NN Q L ++ + DGF Y S + QLTF++
Sbjct: 181 EEKAGQLKVHSTISKGRGRLLSWLAKNNIQELSDIQDFKVDGFEYCTSESTANQLTFIR 240

A related GBS nucleic acid sequence <SEQ ID 10941> which encodes amino acid sequence <SEQ ID 10942> was also identified.

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3399> which encodes the amino acid sequence <SEQ ID 3400>. Analysis of this protein sequence reveals the following:

Possible site: 38
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3759(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 114/242 (47%), Positives = 155/242 (63%)

Query: 1 MIKILIPTAKEMKVQNIAPKLSAQTKIIDIYFSTLTVDLEDIYRINTSAARCEAQRW 60
M+ LIPTAKEM + + L ++ I+ + +T DL YRI +A+ E QRW
45 Sbjct: 1 MLTFLIPTAKEMTIPKESHPHLLPQDSQAILKIMAAMTTEDLAKSYRIKEESAKKEQQRW 60

Query: 61 QDFKAKQLTLNPAIKLFNGLMYRNIKRHNLSSEAQFMENSVFITSALYGIIPAMTLISP 120
QD ++Q PA +LFNGLMYR+IKR L+T E ++ V+ITS+ YGIIEA I+
Sbjct: 61 QDMASQQSLAYPAYQLFNGLMYRHIKRDKLTTQEQAYLTQQVYITSSFYGIIPANHPIAE 120

50 Query: 121 HRLDFNTKIKINNNSLKVFWRENYDTFMQSDDIMVSLLSNEFETVFSPEKQKLIHLNFI 180
HR DF+T+IKI SLK +WR Y+ F + ++SLLS+EF+ VFS +Q I F+
Sbjct: 121 HRHDFHTRIKIEGQSLKSYWRPCYNQFAKEHPQVISLLSSEFDDVFSKDKQLWISPKFM 180

55 Query: 181 EDRDGLKTHSTISKARGKCLTAMMENNCQTLEHLKQLRFDGFCYDNELSDSKQLTFVKQ 242
+++GQ KTHSTISKARG LTA MENNCQT++ LK L F GF Y +LS + ++KK+
Sbjct: 181 AEKEGQFKTHSTISKARGAFLTACMENNCQTVDSLKSLVFAGFYHPDLSTDHEFVYIKKK 242

-1230-

SEQ ID 3398 (GBS428) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 6; MW 30.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 4; MW 55kDa).

GBS428-GST was purified as shown in Figure 220, lane 6-7.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1099

A DNA sequence (GBSx1174) was identified in *S.agalactiae* <SEQ ID 3401> which encodes the amino acid sequence <SEQ ID 3402>. Analysis of this protein sequence reveals the following:

10 Possible site: 23
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.59 Transmembrane 3 - 19 (3 - 19)
 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 10251> which encodes amino acid sequence <SEQ ID 10252> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07024 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 86/275 (31%), Positives = 143/275 (51%), Gaps = 6/275 (2%)
 25 Query: 17 MSYPYKANHSIESITLKVNDLENLVNFYSDIIGLTVIDKSSTRALLGVNQKIPLIILEKT 76
 M + + N ++ + +KV+DL + FY +IIG V+++S A L N + PL+++E+
 Sbjct: 1 MEFHRQPNTFVDLVNIKVSDELRSALTFYQETIIGFQVLERSEERSATLTANGRTPLLVIEQP 60
 30 Query: 77 E---LEKHSTYGLYHTAILVPDEYHLSLALNHLLSQHIPLEGGADHGYSNAIYLSDPEN 133
 + ++ T GLYH A+L+P L LNHL PL+G +DH S AIY +DP+GN
 Sbjct: 61 DPVIAKQPRTTGLYHFALLLPSRADLGRFLNHLLQSGYPLQGASDHLVSEAIYFADPDGN 120
 Query: 134 GIEIYNDKDISMWDIRESGQIIGITERLDIDNLLSLVNPNNYKLSEKTSIGHIHLVK 193
 G+E+Y D+ S WD +G++ TE + +NLL + P L +T +GHIHL V
 35 Sbjct: 121 GVEVYADRPSSWD-WSNGEVKMSTEPHAEENLLAEGKDEPWT-ALPPEITLGHILHLVA 178
 Query: 194 DAKISSKLYQNVFGLDEKFAIPT-ASWIASGNYHHHLAFNNWAGPNLSKNQEDRPGISLL 252
 + + Y G + + A +I++GNYHHH+ N W G E G+
 40 Sbjct: 179 NLFEAETFYIEGLGFNVVARLGNQALFISTGNYHHHIGLNTWNGVGAPTPPEHSVGLKWF 238
 Query: 253 TIAYNDDNLFRLSLKKAQLYQLTFLEKQDHYIIE 287
 ++ Y + + ++ + + K ++I+
 Sbjct: 239 SLITYPSEVRKTVNRLETIGFQVERKHGEEWVID 273

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3403> which encodes the amino acid sequence <SEQ ID 3404>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence
 50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0936(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1231-

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/282 (50%), Positives = 194/282 (68%)

```

5  Query: 17 MSYPYKANHSIESITLKVNDLENLVNFYSDIIGLTVIDKSSTRALLGVNQKIPLIILEKT 76
    M YPY + S+ +++L V DL + FY+ IIGL V+ + +T L + K ++ L +T
Sbjct: 1 MIYPYNSTISLGTVSLNVTDLAKMTTFYTSIIGLQVLSQDTSRQLTTDGKTVILELRQT 60

Query: 77 ELEKHSTYGLYHTAILVPDEYHLSLALNHLSSQHIPLEGGADHGYSNAIYLSDPEGNGIE 136
    L YGLYHTA LVPD + L L LNH L++ I LEG ADHG+S AIYLSDPEGNGIE
10 Sbjct: 61 PLPGDKAYGLYHTAFLVPDRHSLGLVLNHLFLTRISISLEGAADHGHSEAIYLSDPEGNGIE 120

Query: 137 IYNDKDISMWDIRESGQIIGITERLDIDNLLDSLNVNPNYKLSKTSIGHIHLVSVKDAK 196
    IY+DK + WDIR++GQIIG+TE D ++L+ L ++P ++ L++ T I H+HLSVK+A
15 Sbjct: 121 IYHDKAVEHWDIRDNGQIIGVTEPTDTKSILEQLTDIPKHFLAQDTRIRHVHLSVKNAL 180

Query: 197 ISSKLYQNVFGLDEKFAIPTASWIASGNYYHHHLAFNNWAGPNLSKNQEDRPGISLLTIAY 256
    SS LYQ VF L +K IP+ASWIASGNYY+HHLAFN+W+ P L K+QE PG++ LTI
20 Sbjct: 181 ASSLLYQKVFDLGDKMTIPSASWIASGNYYHHHLAFNHWAPYLKKHQEGAPGLAFLTIHI 240

Query: 257 NDDNLFPSATLKKARLHGLAILQEDSSSFTEDEBEGIRVNVIL 298
    LF +LKKA+L+ L L++ + ED +GIR+ V+L
25 Sbjct: 241 ETPLLFPSATLKKARLHGLAILQEDSSSFTEDEBEGIRVNVIL 282

```

SEQ ID 3402 (GBS429) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell
 25 extract is shown in Figure 80 (lane 7; MW 34.2kDa).

GBS429-His was purified as shown in Figure 214, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1100

30 A DNA sequence (GBSx1175) was identified in *S. agalactiae* <SEQ ID 3405> which encodes the amino
 acid sequence <SEQ ID 3406>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

```

35  ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.2362(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

40 A related GBS nucleic acid sequence <SEQ ID 10249> which encodes amino acid sequence <SEQ ID
 10250> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC21682 GB:U32686 conserved hypothetical protein [Haemophilus influenzae Rd]
Identities = 89/261 (34%), Positives = 151/261 (57%), Gaps = 4/261 (1%)
45 Query: 10 MVRLIFSDDIDGTLINSNFVKVTPKTRQGQIKQIVAQGATFVPISARMPEAITPIMEQIGIDS 69
    M + +FSD +GTL+ S ++P+T IK++ A G FVPISAR P I P +Q+ ++
Sbjct: 2 MYKAVFSDFNGLTLLTSQHTISPRTVVVIKRLTANGIPFVPISARSPLGILPYWKQLETNN 61

Query: 70 YIISYNGALIQDMQOKTIASHTMDGQVALQVCSYVSKHYSKIAWNVIYRYHEWYSCDKENE 129
    +++++GALI + + I S ++ + L++ + +++H + N Y ++ ++ D EN+
50 Sbjct: 62 VLVAFSGALILNQLNLEPIYSVQIEPKDILEINTVLAEH-PLLGVNYYTNNDCARDVENK 120

Query: 130 WVQKEEEIVGLQSKEMSLMELEKQDRIHKLKLLMGEPSLMGELENTLKAQYPHLSIAQSAP 189
    WV E + ++ + HK+ ++GE + E+E LK ++PHLSI +S
55 Sbjct: 121 WVIYERSVTIKIEIHPFDEVATRSP---HKIQIIGEABETIEIEVLLKEKFPHLSICRSHA 177

```

-1232-

Query: 190 YFIEIMAPGIEKGSAKTLADYLDISLADSIAGDNYNDLNLLEIVGKGFVMGNAPKDLQ 249
 F+E+M KG + + L DY + + IAFGDN+NDL++LE VG G MGNAP +++
 Sbjct: 178 NFLEVMHKSATKGSAVRFLDYFGVQTNEVIAFGDNFNDLDMLEHVGLGVAMGNAPNEIK 237

Query: 250 ERIGNVTQDNDNDGIYYALVE 270
 + VT N+ DG+ L E
 Sbjct: 238 QAANVVTATNNEGLALILEE 258

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1101

A DNA sequence (GBSx1176) was identified in *S.agalactiae* <SEQ ID 3409> which encodes the amino acid sequence <SEQ ID 3410>. Analysis of this protein sequence reveals the following:

15 Possible site: 19
 >>> May be a lipoprotein

----- Final Results -----

20 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAG07223 GB:AE004801 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 103/283 (36%), Positives = 165/283 (57%), Gaps = 1/283 (0%)

Query: 33 KHIGILQYVEHPSLTATRKGFIKELAKEGYKDGKNIKIEYKNAQCDSNIQSISEKLIKD 92
 K + + VEHP+L A R G + L + GY+DGKN+K +Y++AQG+ I+ K I D
 Sbjct: 31 KSVAVTAIVEHPALDAARDGVKEALQEAGYEDGKNLKWQYQSAQGNLTGTAQIARKFIGD 90

30 Query: 93 NK-LVLGIATPAAQSLTTVSTETPIIFTAVTDPVSAELVKSMKKPEGLATGTSDMSPIKK 151
 +++GIATP+AQ+L + PI+F+ VTDPV A L S + TG SDM + K
 Sbjct: 91 KPDVIVGIATPSAQALVAATKSIPIVFSTVTDVGAHLTPSWEASGNTNVTGVSDMLALDK 150

35 Query: 152 QVSLLRKVMKPKVRVGIMYTTSENRSEVQKQAKKIFQEAGIKTSVKGISSTNDVQDTAK 211
 Q+ L++KV+P KR+G++Y E NS V VK+ K++ + G+ + DV A+
 Sbjct: 151 QIELIKKVVPGAKRIGMVYNPGEANSVVVKELKELLPKMGLSLVEASAPRSDVSSAAR 210

40 Query: 212 SLMSKTEVIFVPTDNIASSVTLLGNLSKELKVPVVGGSADMVPSGLLFSYGADYEALGR 271
 SL+ K + I+ TDN + S+ L + + K+P++ D V G + + G +Y+ +G+
 Sbjct: 211 SLVGKVDATYNTDNNVVSAYEALVKVGNDAKIPLIASDTSVKRGATAALGINYKEMGK 270

Query: 272 QTARQAVKILKGKDVAKVPSEYFQNLKVVVNEDMAKELGIDVS 314
 QT R V+ILKG+ ++ E NL++ VN A++ G+ +S
 45 Sbjct: 271 QTGRMVVRILKGEKPGKIPETSDNLQLFVNPFGAAQKQGVTL 313

There is also homology to SEQ ID 2712.

SEQ ID 3410 (GBS188) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 39 (lane 2; MW 36.6kDa).

- 50 The GBS188-His fusion product was purified (Figure 204, lane 6) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 247), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

-1233-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1102

A DNA sequence (GBSx1177) was identified in *S.agalactiae* <SEQ ID 3411> which encodes the amino acid sequence <SEQ ID 3412>. This protein is predicted to be probable permease of ABC transporter (rbsC). Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have a cleavable N-term signal seq.
10    INTEGRAL    Likelihood = -16.13    Transmembrane    132 - 148 ( 124 - 160)
      INTEGRAL    Likelihood = -6.42    Transmembrane    241 - 257 ( 238 - 258)
      INTEGRAL    Likelihood = -6.32    Transmembrane    264 - 280 ( 260 - 284)
      INTEGRAL    Likelihood = -6.00    Transmembrane    213 - 229 ( 207 - 235)
      INTEGRAL    Likelihood = -4.67    Transmembrane    58 - 74 ( 57 - 75)
15    INTEGRAL    Likelihood = -1.38    Transmembrane    36 - 52 ( 36 - 52)
      INTEGRAL    Likelihood = -0.85    Transmembrane    90 - 106 ( 87 - 106)

----- Final Results -----
      bacterial membrane --- Certainty=0.7453(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAG07224 GB:AE004801 probable permease of ABC transporter
[Pseudomonas aeruginosa]
25    Identities = 114/285 (40%), Positives = 175/285 (61%), Gaps = 3/285 (1%)

Query: 5    ILSGISQGLLWSIMAIGVFITFRILDIALDLAEGAFPMGAAVCALCIVNDINPIVATIAG 64
      + + GL++S+++A+GVFI+FR+L DL+ +G+FP+G AVCA I +P AT+A
30    Sbjct: 6    LFGALEIGLIFSLVALGVFISFRLLRFPDLTVDGSEPLGGAVCATLIALGWDYPYSATLAA 65

Query: 65    MLGGMLAGLVSGFLHTMKIPALLTGIITLTGLYSINLLVLRGSRNVSFALKNTLVTMVTR 124
      G IAGL +G L+ K+KI LL I+ + LYSINL ++G+ NV + TL T++
35    Sbjct: 66    TAAGALAGLATGLLNVLKIMDLLASILMMIALYSINLRIMGKPNVPLIAEPTLFTLLQP 125

Query: 125    LGLNKL SAVLLIGIVCVGLVILILYFLNTQLGLALRATGDNEAMGQANSIKVDRMKMLG 184
      L+ + L+ + V L+L F TQ GLA+RATG N M +A + M +LG
40    Sbjct: 126    EWLSDYVFRPLLVFIVIAAKLLLDWFFTTQKGLAIRATGSRNPRMARAGQGVNTGGMILLG 185

Query: 185    YMIGNGLIALSGALLAQNNGYADLNMGVGTIVIGLASIILAEMIKYLPGLKRLWSIVLG 244
      I N L+AL+GAL AQ G AD++MG+GTIVIGLA++I+ E ++ L +++LG
45    Sbjct: 186    MAISNALVALAGALFAQTQGGADISMGIGTIVIGLAAVIVGESILPSRRLILATLAVILG 245

Query: 245    SVLYRMIIVFILTITD---IDAQMIKLVSAILLALILYVPELRAKL 286
      +++YR I L +D + AQ + LV+A+L+ + L +P ++ +L
50    Sbjct: 246    AIVYRFFIALALNSDFIGLQAQDLNLVTAVLVTVALVIPMMKKRL 290

```

There is also homology to SEQ ID 2716.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1103

A DNA sequence (GBSx1178) was identified in *S.agalactiae* <SEQ ID 3413> which encodes the amino acid sequence <SEQ ID 3414>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```

Possible site: 41
55    >>> Seems to have no N-terminal signal sequence

```

-1234-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3798(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF86640 GB:AF162694 ABC transporter [Enterococcus gallinarum]
 Identities = 171/264 (64%), Positives = 213/264 (79%), Gaps = 1/264 (0%)

10 Query: 3 LLELVNLHKTFEKGTVNENHVLRLDLTIEDGDFISVIGGNGAGKSTLLNCIAGLIPIDQ 62
 +L + +LH+TFEKG+T+NENHVLRLG+DLT+ GDFI++IGGNGAGKSTLLN IAG IP +Q
 Sbjct: 5 VLTISDLHQTFEKGTINENHVLRLGIDLTMSGDFITTIIGGNGAGKSTLLNSIAGTIPTIQ 64

15 Query: 63 GAITLDNQSITKDSVEKRSKDISRVFQDPRMGTAATNLTIENMAIAHKGKGRHIFRQSV 122
 G I L ++ IT+ SV +RSK+ISRVFQDPRMGTA LT+EEN+A+A+KRG R F V
 Sbjct: 65 GKIVLGDKETIRHSVTRRSKEISRVFQDPRMGTAVRLTVEENLALAYKRGQVRG-FSSGV 123

20 Query: 123 TDDDRQLFKKSLSQLGLGLENRMKTDAFLSGGQORQALTLAMATLVRPKLLLLDEHTAAL 182
 R FK+ L++L LGLENR+ T+ LSGGQORQA+TL MATL +PKL+LLDEHTAAL
 Sbjct: 124 KGKHAFFKEKRLARLNLGLENRLTTEIGLLSGGQORQAITLLMATLQPKLILLDEHTAAL 183

25 Query: 183 DPKTSMDVMELTQKVIEEQRILTALMITHNMEHAIAYGNRVLVLYHGKIVVDVKGEAKRNL 242
 DPKTS VM LT ++I+EQ+LTA M+TH+ME AI YGNRL+ML+ GKIVVD+ GE K++L
 Sbjct: 184 DPKTSMTVMALTDQLIQEQQLTAFMVTHDMEDAIRYGNRLIMLHQQGKIVVDITGEEKQSL 243

30 Query: 243 TVAELMELFHKNSGQQLIDDALVL 266
 TV +LM LFH+NSG +L DD L+L
 Sbjct: 244 TVPDLMALFHQNSGTELKDDQLLL 267

There is also homology to SEQ ID 2720:

Identities = 116/249 (46%), Positives = 166/249 (66%), Gaps = 1/249 (0%)

35 Query: 3 LLELVNLHKTFEKGTVNENHVLRLDLTIEDGDFISVIGGNGAGKSTLLNCIAGLIPIDQ 62
 ++EL+N + G + +L + LTI + DF++++GGNGAGKSTL N IAG + + +
 Sbjct: 4 IIEELINATVDVDNGFEDAKTILDNVTLTIYEHDFTILGGNGAGKSTLNFNVIAGTSLR 63

40 Query: 63 GAITLDNQSITKDSVEKRSKDISRVFQDPRMGTAATNLTIENMAIAHKGKGRHIFRQSV 122
 G I + Q +T EKR+ +SRVFQD +MGTA +T+ EN+ IA +RG KR + + +
 Sbjct: 64 GQIRILGQDVTHWPAEKRALYLSRVFQDSKMGTAAPRMTVAENLLIARQGGKRSLSRKI 123

45 Query: 123 TDDDRQLFKKSLSQLGLGLENRMKTDAFLSGGQORQALTLAMATLVRPKLLLLDEHTAAL 182
 T+ F+ + + G GLE ++T A LSGGQORQA+L MATL +P LLLDEHTAAL
 Sbjct: 124 TEHLAS-FEDLVKRTGNGLEKHLTPAGLLSGGQORQALSLLMATLKKPALLLDEHTAAL 182

50 Query: 183 DPKTSMDVMELTQKVIEEQRILTALMITHNMEHAIAYGNRVLVLYHGKIVVDVKGEAKRNL 242
 DPKTS +M+LT + + + LTALMITH+ME A+ YGNRL+++ G I+ D+ K L
 Sbjct: 183 DPKTSQSLMQLTDEFVTKDGLTALMITHMEDALTYGNRLIVMKDGNIIKDLNQMEKEQL 242

55 Query: 243 TVAELMELF 251
 T+ + +LF
 Sbjct: 243 TTIDYYQLF 251

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1104

A DNA sequence (GBSx1179) was identified in *S.galactiae* <SEQ ID 3415> which encodes the amino acid sequence <SEQ ID 3416>. This protein is predicted to be mannose-specific phosphotransferase system component IIAB. Analysis of this protein sequence reveals the following:

60 Possible site: 54

-1235-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3527(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAD46485 GB:AF130465 mannose-specific phosphotransferase system
 component IIAB [Streptococcus salivarius]
 Identities = 287/336 (85%), Positives = 306/336 (90%), Gaps = 6/336 (1%)

Query: 1 MGIGIIIASHGKFAEGIHQSGSMIFGEQEKVQVVTMPNEGPDDLYGHFNNAIAQFDADD 60
 MGIGIIIASHGKFAEGIHQSGSMIFG+QEKVQVVTMP+EGPDDLY HFN+AIAQFDADD

15 Sbjct: 1 MGIGIIIASHGKFAEGIHQSGSMIFGDQEKVQVVTMPSEGPDDLYAHFNDAIAQFDADD 60

Query: 61 EVLVLADLWSGSPFNQASRVNGENPERKMAIITGLNLPMLIQAYTERMMDANAGVEQVAA 120
 E+LVLADLWSGSPFNQASR+ GENP+RK+AIITGLNLPMLIQAYTERMMDANA EQVAA

20 Sbjct: 61 EILVLADLWSGSPFNQASRIAGENPDRKIAIITGLNLPMLIQAYTERMMDANATAEQVAA 120

Query: 121 NIIKESKEGIKALPEELNPVVEATPVAGVPADVPAEVKQSGSIEGTVIGDGKCLKINLAR 180
 NIIKE+K GIKALPEELNP E T A V A P G+IPEGTVIGDGKCLKINLAR

Sbjct: 121 NIIKEAKGGIKALPEELNPAAETT-AAPVEAAAP-----QGAIPEGTVIGDGKCLKINLAR 174

25 Query: 181 IDTRLHGGVATAWTPASKANRIIVASDEVSKDELRLKQLIKQAAPGGVKANVVPISKLE 240
 +DTRLHGGVAT WTPASKA+RIIVASD+V+KDELRLK+LIKQAAP GVKANVVP I KLI+
 Sbjct: 175 LDTRLHGGVATNWTTPASKADRIIVASDDVAKDELRLKELIKQAAPNGVKANVVP I QKLD 234

30 Query: 241 VAKDPRFGNTRALILFETVQDALRAIEGGV EIPELNVGSMASHSTGKTMVNNVLSMDKDDV 300
 +KDPFRFGNT ALILFETVQDALRAIEGGV I ELNVGSMASHSTGKTMVNNVLSMDKDDV
 Sbjct: 235 ASKDPRFGNTHALILFETVQDALRAIEGGVFIKELNVGSMASHSTGKTMVNNVLSMDKDDV 294

Query: 301 AAFEKLRDLGVSFVVRKVPNDAAKKNLFDLINKANVK 336
 A FEKLRDLGV FDVRKVPND+KK+LFDLI KANV+

35 Sbjct: 295 ACFEKLRDLGVFEFVVRKVPNDSKKDLFDLIKANVQ 330

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3417> which encodes the amino acid sequence <SEQ ID 3418>. Analysis of this protein sequence reveals the following:

40 Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3533(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 288/336 (85%), Positives = 308/336 (90%), Gaps = 6/336 (1%)

50 Query: 1 MGIGIIIASHGKFAEGIHQSGSMIFGEQEKVQVVTMPNEGPDDLYGHFNNAIAQFDADD 60
 MGIGIIIASHGKFAEGIHQSGSMIFGEQEKVQVVTMPNEGPDDLYGHFNNAI QFDADD
 Sbjct: 1 MGIGIIIASHGKFAEGIHQSGSMIFGEQEKVQVVTMPNEGPDDLYGHFNNAIQFDADD 60

55 Query: 61 EVLVLADLWSGSPFNQASRVNGENPERKMAIITGLNLPMLIQAYTERMMDANAGVEQVAA 120
 E+LVLADLWSGSPFNQASRV GENP+RKMAIITGLNLPMLIQAYTER+MDA AGVEQVAA
 Sbjct: 61 EILVLADLWSGSPFNQASRVAGENPDRKMAIITGLNLPMLIQAYTERLMDAGAGVEQVAA 120

60 Query: 121 NIIKESKEGIKALPEELNPVVEATPVAGVPADVPAEVKQSGSIEGTVIGDGKCLKINLAR 180
 NIIKESK+GIKALPE+LNPV E V + G+IP GTVIGDGKCLKINLAR
 Sbjct: 121 NIIKESKGIKALPEDLNPVEETAATEKVVNAL-----QGAIPAGTVIGDGKCLKINLAR 174

Query: 181 IDTRLHGGVATAWTPASKANRIIVASDEVSKDELRLKQLIKQAAPGGVKANVVPISKLE 240
 +DTRLHGGVATAWTPASKA+RIIVASDEV++D+LRKQLIKQAAPGGVKANVVPISKLE
 Sbjct: 175 VDTRLHGGVATAWTPASKADRIIVASDEVAQDDLRLKQLIKQAAPGGVKANVVPISKLE 234

-1236-

Query: 241 VAKDPRFGNTRALILFETVQDALRAIEGGVEIPELNVGSMASHSTGKTMVNNVLSMDKDDV 300
 +KDPRFGNT ALILF+T QDALRA+EGGVEI ELNVGSMASHSTGKTMVNNVLSMDK+DV
 Sbjct: 235 ASKDPRFGNTHALILFQTPQDALRAVEGGVEINELNVGSMASHSTGKTMVNNVLSMDKEDV 294

Query: 301 AAFEKLRDLGVSFVVRKVPNDKKNLFDLINKANVK 336
 A FEKLRDLGV+FDVRKVPND+KKNLF+LI K N+K
 Sbjct: 295 ATFEKLRDLGVTFVVRKVPNDSKKNLFBELIQKTNIK 330

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1105

A DNA sequence (GBSx1180) was identified in *S. agalactiae* <SEQ ID 3419> which encodes the amino acid sequence <SEQ ID 3420>. Analysis of this protein sequence reveals the following:

15 Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3873(Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:BAB06625 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 89/267 (33%), Positives = 139/267 (51%), Gaps = 3/267 (1%)

Query: 3 KKIIAVDLDGITLLHNNNTISDYTADTLRKVQAQGHKVIITGRPYRMALAHYLRDLKTP 62
 + +IA+DLDGITLL +N TIS T T++K + GH V+I+TGRPYR ++ +Y L L T
 Sbjct: 4 RHLIALDLDGITLLTDNKTISMKTQTIQKAREAGHIVVISTGRPYRASIQYYQELQLDTA 63

30 Query: 63 MINFNGALTHIPEKKWAFERSATIDKLLLETNLSDAIQADFIASEYRKNFYITMDNRD 122
 ++NFNGA H P+ ++ + + +A I E ++Y+ D
 Sbjct: 64 IVNFNGAFVHHPKDSSFGTYHHPLELSTARQVIETCEAFDVSNI MVEIDYLYLRY--YD 121

35 Query: 123 KINPQLFGVNEITDKMALDVTKITRNPALLMQTRHKDKVELAKELRQHFNHELEVDSWG 182
 ++ Q F + + + K+ +P +L+ + EL L ++ +WG
 Sbjct: 122 BLFIQTFTTEGGQFVEHGNLLKLRDDPTCVLIHPKDDHVBELRSLLDGAHAVIDQRTWG 181

40 Query: 183 GPLNILEFSPKGVNKAYALKHLLKSLNLSQENLIAFGDEHNDTEMLAFAHTGYAMKNANP 242
 P N++E G+NKA LK + + +E +IAFGDE ND EM+ +A G AM NA
 Sbjct: 182 APWNVIEIVKAGMNKAVGLKRIADYYQVPERIIAFGDEDNDFEMIEYAGKGVAMANAID 241

Query: 243 TLLPYADQQIQWINEEDGVAKTLEKLL 269
 L A+ I +NE+DG+A LE+ L
 45 Sbjct: 242 PLKALAN-DITLSNEDDGIAYVLEEAL 267

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3421> which encodes the amino acid sequence <SEQ ID 3422>. Analysis of this protein sequence reveals the following:

50 Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4380(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 188/270 (69%), Positives = 224/270 (82%)

-1237-

Query: 1 MTKKIIAVDLDTLLHNNNTISDYTADTLRKVQAQGHKVIITGRPYRMALAHYLRDLK 60
 MTKK+IA+DLDTLLH++NTIS YT T++ VQ +GH VII+TGRPYRMAL +YL+L+LK
 Sbjct: 1 MTKKLIADLDGTLHHDNTISTYTQKTIKAVQDKGHHVIISTGRPYRMALGYLQLNLK 60

5 Query: 61 TPMINFNGALTHIPEKKWAFERSATIDKKLLLETNLNSDAIQADFIASEYRKNFYITMDN 120
 TP+I FNGALTH+PE+KWA+E + T+DK LL L D Q DFIASEYRKN YITM N
 Sbjct: 61 TPIITFNGALTHMPEQKWAYEHNVTLDKGYLLRLLKYQDDFQMDFIASEYRKNVYITMTN 120

10 Query: 121 RDKINPQLFGVNEITDKMALDVTKITRNPNALLMQTRHKDKYELAKELRQHFNHELEVDS 180
 + I+PQLFGV+EIT MAL++TKITRNPNALLMQT H+DKY LAK +R F E+E+DS
 Sbjct: 121 PESIDPQLFGVDEITQDMALEITKITRNPNALLMQTHHEDKYALAKNMRACFKDEIEIDS 180

15 Query: 181 WGGPLNILEFSPKGVNKAYALKHLLKSLNLSQENLIAFGDEHNDTEMLAFAHTGYAMKNA 240
 WGGPLNILE S K VNKAYAL +LL N+ +++LIAFGDEHNDTEMLAFA TGYAMKNA
 Sbjct: 181 WGGPLNILEISSKNVKNKAYALNYLLGIYNMDKKDLIAFGDEHNDTEMLAFAGTGYAMKNA 240

20 Query: 241 NPTLLPYADQQIQWTNEEDGVAKTLEKLLL 270
 +P LLPYADQQ+ ++NEEDGVAK LE+L L
 Sbjct: 241 SPVLLPYADQQLNFSNEEDGVAKKLEELFL 270

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1106

A DNA sequence (GBSx1181) was identified in *S.agalactiae* <SEQ ID 3423> which encodes the amino acid sequence <SEQ ID 3424>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -7.38	Transmembrane	96 - 112 (90 - 119)
INTEGRAL	Likelihood = -6.58	Transmembrane	28 - 44 (27 - 47)
INTEGRAL	Likelihood = -6.26	Transmembrane	176 - 192 (174 - 193)
INTEGRAL	Likelihood = -5.26	Transmembrane	127 - 143 (126 - 144)
INTEGRAL	Likelihood = -1.59	Transmembrane	4 - 20 (3 - 20)
INTEGRAL	Likelihood = -0.22	Transmembrane	60 - 76 (59 - 78)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1107

45 A DNA sequence (GBSx1182) was identified in *S.agalactiae* <SEQ ID 3425> which encodes the amino acid sequence <SEQ ID 3426>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2025(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

-1238-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1108

- 5 A DNA sequence (GBSx1183) was identified in *S.agalactiae* <SEQ ID 3427> which encodes the amino acid sequence <SEQ ID 3428>. This protein is predicted to be an integral membrane protein. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

10 INTEGRAL Likelihood = -5.41 Transmembrane 180 - 196 (179 - 199)
 INTEGRAL Likelihood = -5.31 Transmembrane 96 - 112 (94 - 114)
 INTEGRAL Likelihood = -2.18 Transmembrane 129 - 145 (129 - 145)
 INTEGRAL Likelihood = -1.33 Transmembrane 37 - 53 (37 - 53)

15 ----- Final Results -----

bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 A related GBS nucleic acid sequence <SEQ ID 8729> which encodes amino acid sequence <SEQ ID 8730> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7

McG: Discrim Score: 5.85

GvH: Signal Score (-7.5): -2.39

25 Possible site: 18

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 4 value: -5.41 threshold: 0.0

30 INTEGRAL Likelihood = -5.41 Transmembrane 176 - 192 (175 - 195)
 INTEGRAL Likelihood = -5.31 Transmembrane 92 - 108 (90 - 110)
 INTEGRAL Likelihood = -2.18 Transmembrane 129 - 145 (129 - 145)
 PERIPHERAL Likelihood = 0.05 57
 modified ALOM score: 1.58

*** Reasoning Step: 3

35 ----- Final Results -----

bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC65028 GB:AE001188 conserved hypothetical integral membrane protein [Treponema pallidum]

Identities = 54/190 (28%), Positives = 93/190 (48%), Gaps = 14/190 (7%)

45

Query: 14 LFFIVISFGIKYYHLQG--PNLIWNMTLALIALDFAYLTSL--FKKKILIGLFALAWFFF 69
 +F +++SFG + L+WN+ LA I + + + F + + L W F
 Sbjct: 3 VFCLLSFGRRCAADNFLSFLVWNLVAFIPWLISAILHVRREFAVRSVQLFLMLLWLLF 62

50

Query: 70 YPNTFYMLTDIIHMHFVGVDVLYNKTNLILYILYVSSILFGFLSGIESFVSVMRKFRISNI 129
 +PN Y+LTDIIH+ L +IL + + + F+S S++ R F I
 Sbjct: 63 FPNAPYIITDIIHLGKGKSFLLYVDLIILLAYSFTGLFYAFVSLHLIESILARDFHIKRP 122

55

Query: 130 FLRWGIIGIVSL-VSSFGIHIGRYARLNSWDILTQPVVINELLAVPSR-----DSFHFI 183
 F II + L + +FGI++GR+ R NSWDI+ + +++++ R D++ F+
 Sbjct: 123 F----IISVFELYLCAPFGIYLGRFLRWNSWDIVLHGRTILSDIGIRVIRPVFYVDTWMFV 178

Query: 184 LGFTFLQVLC 193

-1239-

F + VLC
 Sbjct: 179 FFFGTMLVLC 188

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1109

A DNA sequence (GBSx1184) was identified in *S.agalactiae* <SEQ ID 3429> which encodes the amino acid sequence <SEQ ID 3430>. Analysis of this protein sequence reveals the following:

10 Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -6.79 Transmembrane 171 - 187 (166 - 191)
 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1110

- 25 A DNA sequence (GBSx1185) was identified in *S.agalactiae* <SEQ ID 3431> which encodes the amino acid sequence <SEQ ID 3432>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -10.46 Transmembrane 193 - 209 (191 - 214)
 INTEGRAL Likelihood = -10.30 Transmembrane 99 - 115 (96 - 119)
 30 INTEGRAL Likelihood = -8.17 Transmembrane 454 - 470 (451 - 472)
 INTEGRAL Likelihood = -6.64 Transmembrane 216 - 232 (212 - 236)
 INTEGRAL Likelihood = -6.37 Transmembrane 49 - 65 (43 - 68)
 INTEGRAL Likelihood = -4.88 Transmembrane 362 - 378 (357 - 383)
 INTEGRAL Likelihood = -3.61 Transmembrane 385 - 401 (385 - 402)
 35 INTEGRAL Likelihood = -2.76 Transmembrane 275 - 291 (275 - 291)
 INTEGRAL Likelihood = -1.70 Transmembrane 18 - 34 (18 - 34)
 ----- Final Results -----
 40 bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:AAF95422 GB:AE004299 conserved hypothetical protein [Vibrio cholerae]
 Identities = 193/471 (40%), Positives = 286/471 (59%), Gaps = 42/471 (8%)
 Query: 1 MEKFFKLKEHGTTRTEITAGLTTFMFAMS YILFVNPAILSQTGMPAQGVFLATIIGAVVA 60
 +EK FKL E+GT +RTEI AG+TTF M+YI+FVNPAILS GM VF+AT + A +
 Sbjct: 2 LEKLFKLSEYGTNVRTEILAGVTTPLTMAXYIIFVNPAILSDAGMDRGAVFVATCLAAIG 61
 50 Query: 61 TSVMAFYANLPYAQAPGMGLNAFFTYTVVFALGYTWQEALAMVFICGLISLIITLTKVRK 120
 +M F AN P AQAPGMGLNAFFTY VV +G+TWQ ALA VF G++ ++++L K+R+
 Sbjct: 62 CFIMGFIANPIAQAQAPGMGLNAFFTYGVVLMGHTWQVALAAVFCGVLFIILLFKIRE 121

-1240-

Query: 121 MIESIPTTLKSAITAGIGTFLAYVGIKNAGFLKFSIDPGTYDVVGKGAAGLATITANS 180
 II SIP +L++ I+AGIG FLA++ +KNAG + +P T +V GA L +
 Sbjet: 122 WIINSIPHSRLRTGISAGIGLFLAFIALKNAGIV--VDNPAT--LVSLGAIISLHAV---- 173

5 Query: 181 SATPGLVSFDNPAILLSLIGLSITIFFIVKIRGGIILSILTTLLGILMGVVKLDAINW 240
 L+ +G +TI + +G++G ++++IL T LG++ G V+ I
 Sbjet: 174 -----LAAVGFLTLIGLVYRGVKGAVMIALLAVTALGLVFGDVQWGGIMS 218

10 Query: 241 EATNLSASPRDLKQVFGVALGEKGLISLFSNPSRLPSVLMALAFSLTDIFDTIGTLIGT 300
 +++ +F Q+ A+ E G+IS+ + AF D+PDT GTL+G
 Sbjet: 219 TPPSIAPTF--MQLDFSAVFEIGMISV-----VFAFLFVDLFDTAGTLVGV 262

15 Query: 301 GEKVGLIATTDGNHESKSLDKALYSDLIGTTFGAICGTSNVTYVESAAAGIGAGGRTGLT 360
 K G++ G + L++AL +D T+ GA+ GTSN T+Y+ES +G+ GGRTGLT
 Sbjet: 263 ATKAGLIEKDG---KIPRLNRALLADSTATSVGALLGTSNTTSYIESVSGVAVGRTGLT 319

20 Query: 361 ALVVAGLFAISSFFSPLVSIVPSQATAPILVIVGIMLSNLKDIKWDDMSEAPAFFTSL 420
 A+VV LF ++ FFSPL ++P+ ATA L V I+M+S L I W D++EA P T L
 Sbjet: 320 AVVVGILFLALFFSPLAGMIPAYATAGALFYVAILMMSGLVSDWRDLTEAAPTIVVTCL 379

25 Query: 421 FMGFTYSITYGIAAGFLTYTLAKVIKQAKDIHVVLWILDILFILNFISLA 471
 M T+SI GI+ GF+ Y K+ G+ + + +W++ +F++ +I A
 Sbjet: 380 MMPLTFSIABGISLGFIAAIAIKLFSGKGRSVSLSVWMAAIFVIKYLAA 430

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3433> which encodes the amino acid sequence <SEQ ID 3434>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

30	INTEGRAL	Likelihood = -11.57	Transmembrane	378 - 394 (370 - 419)
	INTEGRAL	Likelihood = -9.29	Transmembrane	202 - 218 (195 - 221)
	INTEGRAL	Likelihood = -7.64	Transmembrane	48 - 64 (46 - 71)
	INTEGRAL	Likelihood = -7.64	Transmembrane	99 - 115 (97 - 118)
	INTEGRAL	Likelihood = -6.90	Transmembrane	225 - 241 (221 - 245)
35	INTEGRAL	Likelihood = -6.05	Transmembrane	468 - 484 (465 - 485)
	INTEGRAL	Likelihood = -4.35	Transmembrane	399 - 415 (395 - 419)
	INTEGRAL	Likelihood = -3.24	Transmembrane	425 - 441 (425 - 442)
	INTEGRAL	Likelihood = -3.08	Transmembrane	18 - 34 (18 - 34)
	INTEGRAL	Likelihood = -2.28	Transmembrane	442 - 458 (442 - 460)
40	INTEGRAL	Likelihood = -0.00	Transmembrane	282 - 298 (282 - 298)

----- Final Results -----
 bacterial membrane --- Certainty=0.5628(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB04327 GB:AP001509 unknown conserved protein [Bacillus halodurans]
 Identities = 192/485 (39%), Positives = 276/485 (56%), Gaps = 53/485 (10%)

50 Query: 1 MEKFFKLSENGTTVSTEIMAGLTTFFAMSYILFVNPSILGAAGMPSNAVFLATIIAAIS 60
 M+++F E+GTT E +AGLTTF +M+YILFVNP ILG AGM AVF+AT +AAAI
 Sbjet: 1 MDRYFGFKEHGTTYGRESIAGLTTFLSMAYILFVNPLILGDAGMDVQAVFMATALAAAI 60

55 Query: 61 TLIMGLFANVPYALAPGMGLNAFFTYTVVFALRFSWQEALAMVFICGLFNIFITVTKFRK 120
 TLIMG+ A P ALAPGMGLNAFF Y+VV + WQ AL VF+ G+ I I'V K R+
 Sbjet: 61 TLIMGILAKYPIALAPGMGLNAFFAYSVVIGMGIDWQLALFGVVFSGIIFILITVFKIRE 120

60 Query: 121 SIIKAIPVSLQHAIGGGIGVFVAYLGFKNANIITFSISAENIVMVNGVEPAKASAKTFAD 180
 II AIP L++A GIG+F+A++G KNA I+
 Sbjet: 121 VIINAIPAELEKNAAGIGLFLAFIQLKNAGIVV----- 154

65 Query: 181 GLLFVDANGGVVPTISSFTDSGVLLAIFGLLLTALVIRNFRGAILIGIVATTLVGIPLG 240
 ++ ++ + LLA FGL++T ++R +G I G++ T +VG+ G
 Sbjet: 155 -----SDEATAVSLGHILNGPTLLACFGLIVTVLFMVRGIQGGIFYGMILTAIVGLISG 208

-1241-

5 Query: 241 IVDVSNLNFGEISHIGRAWTELGTTFPLAAFD-GLSSLFSDSSRLPLVFMFTIFAFSLSDTFD 299
 I+ + I L TF AF+ ++ +FS + + F D FD
 Sbjct: 209 IITYTG-----GGIVSTPPSLAPTFGQAFNIQMADVFSVQ-----FLIVLTLFLVDFD 258

10 Query: 300 TIGTFICTGRRTGIFSQDDENALENSIGFSSKMDRALFADAIGTSIGALVGTSNTTIVVE 359
 T GT G + G F +D++ + +AL AD+ TSIGA++GTS TT Y+E
 Sbjct: 259 TAGTLYGVANQAG-FIKDNK-----LPRAGKALLADSSATSIGAILGTSTTTAYIE 308

15 Query: 420 SKFADALPAFFAAFFMALCYSISYGIAAIFIFYCLVKVVEGKTKDIHPITWGATFLFIVN 479
 +K A+P+F M L YSI+ GIA F+FY + +V+G+ K++HPI++ F+F+
 Sbjct: 369 TKLEIAIPSFLTIVVAMPLTYSIATGIAFGFLFYPTMIVKGRGKEVHPIMYALFFVFLAY 428

20 Query: 480 FIILT 484
 FI L+
 Sbjct: 429 FIFLS 433

An alignment of the GAS and GBS proteins is shown below.

Identities = 258/488 (52%), Positives = 336/488 (67%), Gaps = 17/488 (3%)

25 Query: 1 MEKFFKLKEHGTITRTEITAGLTTFFAMSYILFVNPAISLQTMPPAQGVFLATIIIGAVVA 60
 MEKFFKL E+GTT+ TEI AGLTTFFAMSYILFVNP+IL GMP+ VFLATII A ++
 Sbjct: 1 MEKFFKLSNGTTTVSTEIMAGLTTFFAMSYILFVNPSILGAAGMPNSAVFLATIIAAAI 60

30 Query: 61 TSVMAFYANLPYAQAPGMGLNAFFTYTVVFALGYTWQEALAMVFICGLISLIITLTKVRK 120
 T +M +AN+PYA APMGLNAFFTYTVVFAL ++WQEALAMVFICGL ++ IT+TK RK
 Sbjct: 61 TLIMGLFANVPYALAPGMGLNAFFTYTVVFALRFSWQEALAMVFICGLFNIFITVTKFRK 120

35 Query: 121 MIIESIPTTLKSAITAGIGTFLAYVGIRKAGFLKFSIDPGTYDVV-----GKGA 171
 II++IP +L+ AI GIG F+AY+G KNA + FSI +V K A
 Sbjct: 121 SIIKAIPVSLQHAIGGGIGVFVAYLGFKNANIITFSISAENIVMVNGVEPAKASAKTFAD 180

40 Query: 172 GLATITANSSATPGLVSFDNPAILLSLIGLSITIFFIVKGIRGGIILSILTTTLGILMG 231
 GL + AN P + SF + +LL++ GL +T +++ RG I++ I+ TTL+GI +G
 Sbjct: 181 GLLFVDANGGVVPTISSFTDSGVLLAIFGLLLTTALVIRNFRGAILIGIVATTLVGIPLG 240

45 Query: 232 VVKLDAINWEATNLSASFRDLKQVFGVALGEKGLISLFSNPSRLPSVLMAILAFSLTDIF 291
 +V + +N+ +++ ++ +L F A GL SLFS+ SRLP V M I AFSL+D F
 Sbjct: 241 IVDVSNLNFGEISHIGRAWTELGTTFPLAAF--DGLSSLFSDSSRLPLVFMFTIFAFSLSDTF 298

50 Query: 346 ESAAGIGAGGRTGLTALVVAGLFAISSFFSPLVSIVPSQATAPILVIVGIMMNLKDIK 405
 ESAAGI GGRTGLTA+ A F +S PLV IVP+ ATAP L+IVG+MM+S+ D+
 Sbjct: 359 ESAAGIAGGRTGLTAVSTAVCFLLSILLPLVGVIPAAATAPALIIVGMMVSSFLDVN 418

55 Query: 406 WDDMSEAIPAFFTSLFMGFTYSITYGIAAGFLTITLAKVIKQAKDIHVVLWILDILFIL 465
 W ++A+PAFF + FM YSI+YGIAA F+ Y L KV++G+ KDIH ++W LFI+
 Sbjct: 419 WSKFADALPAFFAAFFMALCYSISYGIAAIFIFYCLVKVVEGKTKDIHPITWGATFLFIV 478

60 Query: 466 NFISLAIL 473
 NFI L IL
 Sbjct: 479 NFIIILTIL 486

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1242-

Example 1111

A DNA sequence (GBSx1186) was identified in *S.agalactiae* <SEQ ID 3435> which encodes the amino acid sequence <SEQ ID 3436>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3221(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB04264 GB:AP001508 unknown conserved protein [Bacillus halodurans]
Identities = 68/147 (46%), Positives = 100/147 (67%), Gaps = 1/147 (0%)
15
Query: 27 MFYTONNEELIALGQKLGTVLKSGDIVLLTGNLKGAGKTTLTGKIAGGLDIKQMIKSPTYT 86
      M  TQ+ E  +A QKL  L +GD++ L G+LGAGKT+ TKG+A GL IK+++KSPT+T
Sbjct: 5  MMITQSPEATMAFAQKLADKLLAGDVITILEGDLGAGKTSFTKGLALGLGIKRVVKSPTFT 64

20
Query: 87 IVREYEGRVFLYHLDVYRIGDDPDSIDLDDFLFGQGVTVIEWGELLSDNLINNYLEIVIT 146
      I+REY+GR+PLYH+DVYR+ ++ + + D++ G GVTV+EW L+ L  L I IT
Sbjct: 65 IIREYKGRFLYHMDVYRLNEEEEDLGDFEYFHDGVTVVEWASLIEGRLPVRLAITIT 124

Query: 147 RSNQG-RQVQLEAYGHRAREIIEAIQD 172
25     + + RQ+  AYG R  E+++ + D
Sbjct: 125 HAGENERQLSFTAYGERWEEVLKELLID 151

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3437> which encodes the amino acid sequence <SEQ ID 3438>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 43
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1202(Affirmative) < succ>
35     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 97/142 (68%), Positives = 122/142 (85%)
40
Query: 27 MFYTONNEELIALGQKLGTVLKSGDIVLLTGNLKGAGKTTLTGKIAGGLDIKQMIKSPTYT 86
      MFY++NE  L A G+ LGT L  GD+++L+G+LGAGKTTL KGIAG+ I QMIKSPTYT
Sbjct: 1  MFYSENEYTLKAYGETLGTYSIGDVIVLSGDLGAGKTTLAKGIAGKMGISQMIKSPTYT 60

45
Query: 87 IVREYEGRVFLYHLDVYRIGDDPDSIDLDDFLFGQGVTVIEWGELLSDNLINNYLEIVIT 146
      IVREYEGR+PLYHLD+YR+GDDPDSIDLDDFLFG GVTVIEWGELL + L+ +YL+I IT
Sbjct: 61 IVREYEGRLPLYHLDIYRVGDDPDSIDLDDFLFGNGVTVIEWGELLGEGLLQDYLQITTT 120

Query: 147 RSNQGRQVQLEAYGHRAREIIE 168
50     + ++GRQ+ L A+G R+R+++E
Sbjct: 121 KRDKGRQLDLLAHGERSRQLLE 142

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1112

A DNA sequence (GBSx1187) was identified in *S.agalactiae* <SEQ ID 3439> which encodes the amino acid sequence <SEQ ID 3440>. Analysis of this protein sequence reveals the following:

-1243-

Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1782(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAD35662 GB:AE001732 conserved hypothetical protein [Thermotoga maritima]
 Identities = 56/163 (34%), Positives = 94/163 (57%), Gaps = 1/163 (0%)

Query: 24 EASREEASAIIEFLNTVTEETDFILHTVSNQLSLSEMETFIENTLMTKNCICLIAKLKKNK 83
 EAS +A I+E+L VT ETDF++ +S +I + ++ ++ +
15 Sbjct: 18 EASIWDARRIVEYLKEVTSETDFLITRPDEVYDVSTERNYIRMYRSNPGKLMIVGEINRE 77

Query: 84 VIGLITIISQSDIEIEHVGDLFIAVQKDYWGYGIGHILMEEAIEWASDNDITRRLSVQ 143
 ++ L+T +HVG++ I+V+K YW GIG ++ AIEWA N R++L V
20 Sbjct: 78 IVSLLTFTGFGRKRTKHVGEIGISVKKRYWNIGICTRMITSAIEWARRNGFI-RIQLEVL 136

Query: 144 GRNERAIHLYQKFGFEIDGLQTRGIKRENGEFLDIYRMSKLID 186
 NERAI LY+K GFE++G++ + ++R++G F D+ M+ L+D
25 Sbjct: 137 KSNERAISLYRKLGFELEGIRKKA VRDDGSFEDVLVMALLLD 179

25 There is also homology to SEQ ID 1724.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1113

30 A DNA sequence (GBSx1188) was identified in *S.agalactiae* <SEQ ID 3441> which encodes the amino acid sequence <SEQ ID 3442>. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

35 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB15582 GB:Z99122 membrane-bound protein [Bacillus subtilis]
 Identities = 108/324 (33%), Positives = 178/324 (54%), Gaps = 33/324 (10%)

Query: 5 KKITLMFSAIILTIVIALGV--YVASAYNFSTNELSKTFKDFKLAKS--KSHAIEETKPF 60
 KK TL+ + + + ++ LG Y ++ + + ++ + +K K +I + PF
45 Sbjct: 8 KKKTLLLTILTITIGLLVLGTGGYAYYLWHKAASTVASIHESIDKSKKRDKEVSINKKDPF 67

Query: 61 SILLMGVDTGSEHRKSKWSGNSDSMILVTINPKTNKTTMTSLERDVLIKLSGPKNNQQTG 120
 S+L+MGVD + G +D++I +T+NPKTN T M S+ RD K+ G G
50 Sbjct: 68 SVLINGVDERDGDK-----GRADTLIYMTVNPKNNTTDMVSI PRDITYTKIIGK-----G 116

Query: 121 VEAKLNAAYASGGAEMALMTVQDLLEDINVDYFMQINMQGLVDLVNAVGGITVINKFDFPI 180
 K+N +YA GG +M + TV++ LD+ VDYP++NM+ D+V+ +GGITV + F F
55 Sbjct: 117 TMDKINHSAFAGGTQMTVDTVENFLDVPVDYFVKVNMESEFRD VVDITLGGITVNSTFAFSY 176

Query: 181 SIAANEPEYKAVVEPGTHKINGEQALVYSRMRYDDPEGDYGRQKRQREVIQVKKILAL 240
 + G +NG++AL Y+RMR +DP GD+GRQ RQR+VIQ ++ K +
Sbjct: 177 DGYS-----FGKGEITLNGKEALAYTRMRKEDPRGDFGRQDRQROVIQGIINKGANI 228

Query: 241 NSISSYKKILSAVSNMQTNIEISSKTIPNL----LAYKDSLEHIKSYQLKGEDATLSGD 296

-1244-

+SI+ + + V NN++TN+ T N+ YK + +HIK ++LKG T +G
 Sbjct: 229 SSITKFGDMFKVVENNVKTNL-----TFDNMWDIQSDYKGARKHIKQHELKG-TGTKING 282

Query: 297 GSYQILTCKHLLAVQNRIKKELDK 320

Y + L + +K+ L+K

Sbjct: 283 IYYYQADESALSDITKELKESLEK 306

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2763> which encodes the amino acid sequence <SEQ ID 2764>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 288/436 (66%), Positives = 342/436 (78%), Gaps = 22/436 (5%)

Query: 1 MKIWKKITLMFSAIILTTVIALGVVVASAYNFSSTNELSKTFKDFKLAKSKSHAIEETKPF 60

MKI KKI LMF+AI+LTTV+ALGVY+ SAY FST ELSKTFKDF + +KS AI++T+ F

Sbjct: 1 MKIGKKIVLMFTAIVLTTVLALGVYITSAYTFSTGELSKTFKDFSTSSNKSDAIKQTRAF 60

Query: 61 SILLMGVDTGSEHRKSKWSGNSDSMILVTINPKTKTMTSLERDVLIKLSGPKNNQQTG 120

SILLMGVDTGS R SKW GNSDSMILVT+NPKT KTTMTSLERD L LSGPKNN G

Sbjct: 61 SILLMGVDTGSSERASKWEGNSDSMILVTNPKTKTMTSLERDTLTLTSGPKNNEMNG 120

Query: 121 VEAKLNAAAYASGCAEMALMTVQDLLDINVDYFMQINMQGLVDLVNAVGGITVTNKFDFPI 180

VEAKLNAAAYA+GGA+MA+MTVQDLL+I +D ++QINMQGL+DLVNAVGGITVTN+FDFFPI

Sbjct: 121 VEAKLNAAAYAAGGAQMAIMTVQDLLNITIDNYVQINMQGLIDLNAVGGITVTNEFDFFPI 180

Query: 181 SIAANEPEYKAVVEPGTHKINGEQALVYSRMRYDDPEGDYGRQKRQREVIQKVLKKILAL 240

SIA NEPEY+A V PGTHKINGEQALVY+RMRYDDPEGDYGRQKRQREVIQKVLKKILAL

Sbjct: 181 SIAANEPEYQATVAPGTHKINGEQALVYARMRYDDPEGDYGRQKRQREVIQKVLKKILAL 240

Query: 241 NSISSYKILSAVSSNMQTNIEISSKTI PNLLAYKDSLEHIKSYQLKGEDATLSDGGSYQ 300

+SISSY+KILSAVS+NMQTNIEISS+TIP+LL Y+D+L IK+YQLKGEDATLSDGGSYQ

Sbjct: 241 DSISSYKILSAVSSNMQTNIEISSRTIPSLLYGRDALRTIKTYQLKGEDATLSDGGSYQ 300

Query: 301 ILTKKHLLAVQNRIKKELDKRKTLSAILEYEDYGTASNDSSSTYSSTQENNYNTT- 359

I+T HLL +QNRI+ EL + LKT+A +YE+ YG ST S T NNY+++

Sbjct: 301 IVTSNHLLEIQNRIRTELGLHKVNQLKTNATVYENLYG-----STKSQTVNNYDSSG 353

Query: 360 ---PYSEAPPSYSG-----NTIYSSETNQTHQNYNSSTPASNYSSNTNTGQADSSGSV 411

YS++ SY+ +T S+ T+Q + + + +TP+S+ S ++ SSGS

Sbjct: 354 QAPSYSDSHSSYANYSSGVDTGQSASTDQDSTASSHRPATPSSS-SDALAADESSSSGS- 411

Query: 412 NNHNGAATPNPNTGTQ 427

G+ P N Q

Sbjct: 412 ----GSLVPPANINPQ 423

SEQ ID 3442 (GBS54) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 8; MW 48.4kDa).

The GBS54-His fusion product was purified (Figure 98A; see also Figure 194, lane 6) and used to immunise mice (lane 1+2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 98B), FACS (Figure 98C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS and that it is an effective protective immunogen.

-1245-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1114

A DNA sequence (GBSx1189) was identified in *S.agalactiae* <SEQ ID 3443> which encodes the amino acid sequence <SEQ ID 3444>. This protein is predicted to be Vesl-1L. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -1.44    Transmembrane    3 - 19 ( 3 - 19)
----- Final Results -----
      bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3445> which encodes the amino acid sequence <SEQ ID 3446>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
>>> Seems to have a cleavable N-term signal seq.
----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 42/98 (42%), Positives = 64/98 (64%)

Query: 1  MKIGRLIALGLVSLGALELYKNRKTIKDSYQNTKNETDSAKLKLERIKNDLAIISQEKEK 60
          MK+ +IA+GL+S A + Y+ R TIK+ ++ D+A+L L+ IK +L +I + +
Sbjct: 1  MKVKTFVIAVGLLSFTAYKAYQKRCITIKELLSISRQAKDAAQLDLNLIKANLDLIHSQGKV 60

Query: 61  IRLISQELNHKFOVFNKDIQPRLEEINQRMAYQEKEDE 98
          I+ ISQ+L HK++ FN++ Q L EI RMAKYQE E
Sbjct: 61  IQNISQDLAHKWRYPNQETQAHLTEIQNRMAKYQEDSE 98

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1115

A DNA sequence (GBSx1190) was identified in *S.agalactiae* <SEQ ID 3447> which encodes the amino acid sequence <SEQ ID 3448>. This protein is predicted to be Hit-like protein involved in cell-cycle regulation (hit). Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2694(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-1246-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04908 GB:AP001511 Hit-like protein involved in cell-cycle
regulation [Bacillus halodurans]
Identities = 74/137 (54%), Positives = 95/137 (69%), Gaps = 2/137 (1%)

5 Query: 3 NCIFCKIISGEIPSSKVYEDDEVLAFLDITQTTTGHSTLLIPKKHVRNVLEMDEKTAQITF 62
NCIFCKII+GEIPS+ VYEDD V AFLDI+Q T GHSTL+IPK H RNV E+ E+ A F
Sbjct: 6 NCIFCKIIAGEIPSATVYEDDHVYAFLDISQVTKGHSTLVIPKVKHVRNVFELSEEIASSLF 65

10 Query: 63 ERLPKVARAVQAATKAKGMNIINNNEEIAAGQTVFHAHVHLVPRFDES DGIKIHYTTHEPD 122
+PK++RA+ A + GMNI+NNN E AGQTVFH H+HL+PR+ E DG + H
Sbjct: 66 AAVPKISRINDAFQPTGMNIVNNNGEAGQTVFHYHLHLPRYGE GEGYGA VWKDHSSQ 125

15 Query: 123 F--EALAKLAKEIRKEI 137
+ + L L+ IR+ +
Sbjct: 126 YSGDDLQVLSSSIREHL 142

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3449> which encodes the amino acid sequence <SEQ ID 3450>. Analysis of this protein sequence reveals the following:

20 Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0125 (Affirmative) < succ>
25 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 97/137 (70%), Positives = 117/137 (84%)

30 Query: 1 MDNCIFCKIISGEIPSSKVYEDDEVLAFLDITQTTTGHSTLLIPKKHVRNVLEMDEKTAQI 60
M+NCIFC II G+IPSSKVYED++VLAFLDI+QTT GHSTL+IPK+HVRN+LEM +TA
Sbjct: 1 MENCIFCSIIQGDIIPSSKVYEDDEVLAFLDISQVTKGHSTLVIPKQHVRNLLLEMTAETASH 60

35 Query: 61 TFERLPKVARAVQAATKAKGMNIINNNEEIAAGQTVFHAHVHLVPRFDES DGIKIHYTTHE 120
F R+PK+ARA+Q+AT A MNIINNNE +AGQTVFHAHVHLVPR++E DGI I YTTHE
Sbjct: 61 LFARIPKIARAIQSATGATAMNIINNNEALAGQTVFHAHVHLVPRYNEEDGISIQYTTHE 120

40 Query: 121 PDFEALAKLAKEIRKEI 137
PDF L KLA++I +E+
Sbjct: 121 PDFPVLEKLARQINQEV 137

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 1116

A DNA sequence (GBSx1191) was identified in *S.agalactiae* <SEQ ID 3451> which encodes the amino acid sequence <SEQ ID 3452>. Analysis of this protein sequence reveals the following:

50 Possible site: 36
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
55 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10923> which encodes amino acid sequence <SEQ ID 10924> was also identified.

-1247-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3452 (GBS87) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 8 (lane 3; MW 19.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 10; MW 44kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1117

A DNA sequence (GBSx1192) was identified in *S.galactiae* <SEQ ID 3453> which encodes the amino acid sequence <SEQ ID 3454>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -6.53    Transmembrane  143 - 159 ( 141 - 161)
----- Final Results -----
          bacterial membrane --- Certainty=0.3612(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9563> which encodes amino acid sequence <SEQ ID 9564> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12844 GB:Z99109 ABC transporter (ATP-binding protein)
[Bacillus subtilis]
Identities = 137/242 (56%), Positives = 181/242 (74%)

Query: 1  MTMLKIENVTTGGYVNIPVLKNISFEVNDGELVGLIGLNGAGKSTTINEIIGILRPYQGGDI 60
      M++L ++++TGGY  PVLKN+SF +  ++VGLIGLNGAGKSTTI  IIG++ P++G I
Sbjct: 1  MSLLSVKDLTGGYTRNPVLKNVSFTLEPNQIVGLIGLNGAGKSTTIRHIIGLMDPHKGS I 60

Query: 61  TIDGISLEADQELYRKIGFIPETPSLYEELTLREHLETVAMAYDIATDEV MARAQK LLE 120
      ++G +  D E YR +  +IPETP LYEELTL EHLE  AMAY ++ + + R  LL+
Sbjct: 61  ELNGKTF AEDPEGYRSQFTYIPETPVLYEELTLMEHLELTAMAYGLSKETMEKRLPPLK 120

Query: 121 MFRITDKLDWFPMHFSKGMKQKVMII CAFVVS PSLFIVDEPFLGLDPLAISDLINLLAE 180
      FR+  +L WFP HFSKGMKQKVM I+CAF+  P+L+I+DEPFLGLDPLAI+ L+  + E
Sbjct: 121 EFRMEKRLKWFP AHFSKGMKQKVMIMCAFLAEPALYIIDEPFLGLDPLAINALLERMNEA 180

Query: 181 KAKGKSILMSTHVLDSAEKMCDFVILHKGEIRAVGTLEELRAIFGDSNANLNDIYIALT 240
      K G S+LMSTH+L +AE+ CD F+ILH GE+RA GTL ELR  FG  +A L+D+Y+ LT
Sbjct: 181 KKGASVLMSTHILATAERYCDSFIILHNGEVRARGTLSELREQFGMKDAALDDLYLELT 240

Query: 241 KE 242
      KE
Sbjct: 241 KE 242
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3455> which encodes the amino acid sequence <SEQ ID 3456>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -5.04    Transmembrane  141 - 157 ( 139 - 158)
```

-1248-

----- Final Results -----

5 bacterial membrane --- Certainty=0.3017(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB12844 GB:Z99109 ABC transporter (ATP-binding protein)
 [Bacillus subtilis]
 10 Identities = 139/241 (57%), Positives = 189/241 (77%)

 Query: 1 MLNLIKNTGGYHNIPVLNDVSFSDNGELVGLIGLNGAGKSTTINEIIGFLKPYQGSISI 60
 +L++K+LTGGY PVL +VSF+++ ++VGLIGLNGAGKSTTI IIG + P++GSI +
 15 Sbjct: 3 LLSVKDLTGGYTRNPFVLKNVSFTLEPNQIVGLIGLNGAGKSTTIRHIIGLMDPHKGSIEL 62

 Query: 61 DGLTLAENAVAYRQKIGFIPETPSLYEELTLSEHINTVAMAYDIDLEVAQKRAQPFLMEF 120
 +G T AE+ YR + +IPETP LYEELTL EH+ AMAY + E +KR P L+ F
 20 Sbjct: 63 NGKTFAEIDPEGYRSQFTYIPETPVLVEELTLMEHLELTAMAYGLSKETMEKRLPPILLKEF 122

 Query: 121 RLTDKLEWFPVNFSGMKQKVMIICAFVIDPSLFIIDEPFLGLDPLAISDLIQTLEVEKA 180
 R+ +L+WFP +FSKGMKQKVMIICAF+ +P+L+I+DEPFLGLDPLAI+ L++ + K
 25 Sbjct: 123 RMEKRLKWFPFAHFSKGMKQKVMIMCAFLAEPALYIIDEPFLGLDPLAINALLERMNEAKK 182

 Query: 181 KGKSILMSTHVLDSAEKMDRFRVILHHGQVRAQGTLDLQAFGDRSASLNDIYLAITKED 241
 G S+LMSTH+L +AER CD F+ILH+G+VRA+GTL++L+E FG + A+L+D+YL LTKED
 30 Sbjct: 183 GCASVLMSTHILATAERYCDSFTILHNGEVRARGTSELREQFGMKDAALDDLYLELTKEK 243

An alignment of the GAS and GBS proteins is shown below.

Identities = 181/240 (75%), Positives = 208/240 (86%)
 30 Query: 3 MLKIENVTTGGYVNIPVLKNISFEVNDGELVGLIGLNGAGKSTTINEIIGILRPYQGDITI 62
 ML I+N+TGGY NIPVL ++SF V++GELVGLIGLNGAGKSTTINEIIG L+PYQG I+I
 Sbjct: 1 MLNLIKNTGGYHNIPVLNDVSFSDNGELVGLIGLNGAGKSTTINEIIGFLKPYQGSISI 60

 35 Query: 63 DGISLEADQELYRKKIGFIPETPSLYEELTLREHLETVAMAYDIATDEVMAKRAQKLEMF 122
 DG++L + YR+KIGFIPETPSLYEELTL EH+ TVAMAYDI + RAQ LEMF
 Sbjct: 61 DGLTLAENAVAYRQKIGFIPETPSLYEELTLSEHINTVAMAYDIDLEVAQKRAQPFLMEF 120

 40 Query: 123 RLTDKLDWFPVNFSGMKQKVMIICAFVVSFSLFIVDEPFLGLDPLAISDLINLLAEKA 182
 RLTDKL+WFP++FSKGMKQKVMIICAFV+ PSLFI+DEPFLGLDPLAISDLI L EKA
 Sbjct: 121 RLTDKLEWFPVNFSGMKQKVMIICAFVIDPSLFIIDEPFLGLDPLAISDLIQTLEVEKA 180

 45 Query: 183 KGKSILMSTHVLDSAEKMDRFRVILHHGQVRAQGTLDLQAFGDRSASLNDIYLAITKE 242
 KGKSILMSTHVLDSAE+MCDRFRVILH G++RA GTL +L+ FGD +A+LNDIY+ALTKE
 Sbjct: 181 KGKSILMSTHVLDSAEKMDRFRVILHHGQVRAQGTLDLQAFGDRSASLNDIYLAITKE 240

SEQ ID 3454 (GBS353) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 2; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 6; MW 55kDa).

50 GBS353-GST was purified as shown in Figure 216, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1118

55 A DNA sequence (GBSx1193) was identified in *S.agalactiae* <SEQ ID 3457> which encodes the amino acid sequence <SEQ ID 3458>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

-1249-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1475 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1119

A DNA sequence (GBSx1194) was identified in *S.galactiae* <SEQ ID 3459> which encodes the amino acid sequence <SEQ ID 3460>. Analysis of this protein sequence reveals the following:

Possible site: 44

15 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.68	Transmembrane	57 - 73 (50 - 80)
INTEGRAL	Likelihood = -8.49	Transmembrane	122 - 138 (103 - 152)
INTEGRAL	Likelihood = -6.58	Transmembrane	319 - 335 (308 - 337)
INTEGRAL	Likelihood = -4.99	Transmembrane	252 - 268 (249 - 273)
20 INTEGRAL	Likelihood = -4.19	Transmembrane	104 - 120 (103 - 121)
INTEGRAL	Likelihood = -3.50	Transmembrane	231 - 247 (229 - 248)
INTEGRAL	Likelihood = -1.91	Transmembrane	298 - 314 (298 - 314)
INTEGRAL	Likelihood = -1.44	Transmembrane	28 - 44 (27 - 44)

25 ----- Final Results -----

bacterial membrane --- Certainty=0.6074 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12845 GB:Z99109 ABC transporter (membrane protein) [Bacillus subtilis]
 Identities = 101/409 (24%), Positives = 187/409 (45%), Gaps = 76/409 (18%)

35 Query: 1 MKKLEFNKRRSLFLTONSKYLRYVFNDFHVLVLMFLSGFLLYQYSQLLKDFPKTHWPIIVI 60
 M ++ R ++ Y++Y+ NDH V+VL+ F YS+ ++D P H+P +
 Sbjct: 4 MLDIWQSRILQEHKETRTYMKYMLNDHLVIVLIFFLAGAASWYSKWIRDIP-AHFPSFWV 62

40 Query: 61 VSIILMLLAMGSIASYLEPADKQFLIKEEAIKEIINSAKKRTYI----- 106
 ++++ ++L + + L+ AD FLL E ++ + A +Y+
 Sbjct: 63 MAVLPSLVLTSSVVRTLLKEADLVFLLPLEAKMBEPYLKQAFVYSYVSQFLPLIALSIVAM 122

45 Query: 107 --FWLVIQTLFLVLISPIILKGL----- 128
 ++ V LV + + ++L L
 Sbjct: 123 PLYFAVTPGASLVSYAAVVFQLLLLKAWNQVMWRTTFQNDRSMDVIRFAANTLVL 182

50 Query: 129 -----SVFMITLLIFGLGIKWLIVITYKVKVFYNNQNLNWDAAINHEQERKQSILKFFSL 183
 SV+M LL++ + + +L++ K + W++ I E RKQ + +L
 Sbjct: 183 YFVFQSVYMYALLVYVIMAVLYLYMSSAAK----RKTFFKWSHIESELRRKQRFYRIANL 238

55 Query: 184 FTNVKGISTSVKRRSFLDGILKLSKTPSRLWTNLFVRAFLRSSDYLGTLIRLVTLNLS 243
 FT+V + KRR++LD +L+L+ + + +F RAFLRSSDYLG+ +RL + L
 Sbjct: 239 FTDVPHLRKQAKRRAYLDLFLRLVPFEQRTTFAYMFTRAFLRSSDYLGILVRLTIVFALI 298

Query: 244 VIFVNETYLALALAFVFN-YLLLQLLALGHFDYQYMNQLYPVRNLAKASQLKGFLRLV 302
 +++V+ + L A+ VF ++ QLL L HFD+ + +LYPV+ K ++LK + +L
 Sbjct: 299 IMYVSASPLIAAVLTVFAIFITIGIQLPLFGHFDHLALQELYPVQ---KETKLKSYFSLL 355

Query: 303 SYAVTVIDSI-----LIRELKPVILLIVLMLIVTEYYIPYKIKK 341
 A+++ + L L +I VL+ +V Y+ ++KK

-1250-

Sbjct: 356 KTALSIQALLMSVASAYAAGLTGFLYALIGSAVLIFVVLPAYMTTRLKK 404

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3461> which encodes the amino acid sequence <SEQ ID 3462>. Analysis of this protein sequence reveals the following:

5 Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -14.91 Transmembrane 126 - 142 (119 - 151)
 INTEGRAL Likelihood = -9.77 Transmembrane 320 - 336 (311 - 339)
 INTEGRAL Likelihood = -6.37 Transmembrane 59 - 75 (53 - 79)
 10 INTEGRAL Likelihood = -4.94 Transmembrane 28 - 44 (22 - 47)
 INTEGRAL Likelihood = -4.73 Transmembrane 250 - 266 (249 - 273)
 INTEGRAL Likelihood = -4.04 Transmembrane 231 - 247 (229 - 248)
 INTEGRAL Likelihood = -3.19 Transmembrane 298 - 314 (295 - 315)
 15 INTEGRAL Likelihood = -2.28 Transmembrane 103 - 119 (103 - 119)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.6965(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB12845 GB:Z99109 ABC transporter (membrane protein) [Bacillus subtilis]
 Identities = 96/403 (23%), Positives = 173/403 (42%), Gaps = 78/403 (19%)
 25 Query: 1 MKALFLKRRQDFQKQNKYLRVVLNDHFVLVLMFLGFMVQYQQLLN----HFPT---- 52
 M ++ R Q+ K+ Y++Y+LNDH V+VL+ F L A Y + + HFP+
 Sbjct: 4 MLDIWQSRLQEHKETRTYMKYMLNDHLVIVLIFFLAGAASWYSKWIRDIPAHFSPFWVM 63
 30 Query: 53 -----NHLPIQVCLGILIPLLLSM----- 71
 L + L L+PL M
 Sbjct: 64 AVLFSLVLTSSYVRTLLKEADLVFLLPLEAKMEPYLKQAFVYSYVSQLFPLIALSIVAMP 123
 35 Query: 72 -----GSIATYLEEADQHFLLPKEEIVISYI-----KQERLSFLLWGTLTQTAVLL 117
 S+ +Y Q LL +V+ + + +R+ ++ T VL
 Sbjct: 124 LYFAVTPGASLVSYAAVVFQLLLLKAWNQVMENRTTFQNDRSMKRMVDVIRFAANTLVLY 183
 40 Query: 118 FLYPIFRRLGLSLFIFIIIVLILLALKRVVLSRKTRYFLRGNRLDWAKAVAFESNRKQSI 177
 F++ S++++ +LV +++A+ + +S + W + E RKQ
 Sbjct: 184 FVFQ-----SVVYALLVYVIMAVLYLYMSSAAKR----KTFKWESHIESELRRKQRF 232
 45 Query: 178 LKFSYSLFTTVKGIKTKVKTERTYLNPLKLVKQTPSNLWLSLYARAFLRSSDYLGFLRLM 237
 + +LFT V + + K R YL+ LL+LV + ++ RAFLRSSDYLG+ +RL
 Sbjct: 233 YRIANLFTDVPHLRKQAKRRAYLDLFLRLVPFEQRKTFAYMFTRAFLRSSDYLGILVRLT 292
 50 Query: 238 LLSSLSVFFIHNLYLSVSLALIFN-YLVVFQLLSLYHYDYHYMTSLYPENSRSKKKNML 296
 ++ +L + ++ L ++ +F ++ QLL L+ H+D+ + LYP +K K+
 Sbjct: 293 IVFALIIMYVSASPLIAAVLTVFAIFITGIQLPLFGHFDHLALQELYPVQKETKLKSYF 352
 55 Query: 297 SFLR-GLSFLMLIVNMLCCSSAPKA--LILIVGMVFIACIYLP 336
 S L+ LS L++++ +A L ++G + + LP
 Sbjct: 353 SLKKTALSIQALLMSVASAYAAGLTGFLYALIGSAVLIFVVL 395

An alignment of the GAS and GBS proteins is shown below.

Identities = 170/344 (49%), Positives = 237/344 (68%)
 55 Query: 1 MKKLFNKRSLFLTQNSKYLRVFNDFVFLVLMFLSGFLLYQYSQLLKDFPKTHWPIIVI 60
 MK LF KRR F Q +KYLRYV NDHFVLVLMFL GF + QY QLL FP H PI V
 Sbjct: 1 MKALFLKRRQDFQKQNKYLRVVLNDHFVLVLMFLGFMVQYQQLLNHFPTNHLPIQVC 60
 60 Query: 61 VSIILMLLAMGGIASYLEPADKQFLLIKEBAIKEIINSAKRTYIFWLVIQTLFLVLIS 120
 + I+I +LL+MG IA+YLE AD+ FLL KEE + I A++ +++ W +QT L+ +
 Sbjct: 61 LGILIPLLSMGSIATYLEEADQHFLLPKEEIVISYIKQERLSFLLWGTLTQTAVLLFLY 120
 Query: 121 PILIKLGLSVFMITLLIFGLGIKWLVTYKVKVFYNNQNLNWDAAINHEQERKQSILKF 180

-1251-

PI +LGLS+F+ +L+ L +K +V++ K + F L+W A+ E RKQSILKF
 Sbjct: 121 PIFRRLGLSLFIFIFILVLILLALKRVVLSRKTRYFLRGNRLDWAKAVAFESNRKQSILKF 180
 Query: 181 FSLFTNVKGISTSVKRRSFLDGILKLISKTPSRLWTNLFVRAFLRSSDYLGTLTIRLVTLN 240
 +SLFT VKGIST VK R++L+ +LKL+ +TPS LW +L+ RAFLRSSDYLGTL +RL+ L+
 Sbjct: 181 YSLFTTVKGISTKVKERTYLNPLLVKQTPSNLWLSLYARAFLRSSDYLGTLFLRLMLLS 240
 Query: 241 ILSVIFVNETYLALALAFVFNLYLLLFQLLALGHFDYQYMNQLYPVRLNAKASQLKGFLR 300
 LSV F++ YL+++LA +FNYL++FQLL+L +H+DY YM LYP +K + FLR
 Sbjct: 241 SLSVFFIHNLYLSVSLALIFNYLVVFQLLSLYYHYDYHYMTSLYPENSRSKKKNMLSFLR 300
 Query: 301 VLSYAVTVIDSILIRELKPVILLIVMLIVTEYYIPYKIKKID 344
 LS+ + +++ + ++LIV M+ + Y+PYK+KK+ID
 Sbjct: 301 GLSFLMLIVNMLCCSSAPKALILIVGMVFIACIYLPYKLKKIID 344

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1120

A DNA sequence (GBSx1195) was identified in *S.agalactiae* <SEQ ID 3463> which encodes the amino acid sequence <SEQ ID 3464>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2821(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00284 GB:AF008220 YtmP [Bacillus subtilis]
 Identities = 69/214 (32%), Positives = 121/214 (56%), Gaps = 1/214 (0%)
 Query: 12 PLRGKSGKAYIGTYPNGERVFVKYNTTPIPLPALAKEQIAPQLLWARRTSNGDMMSAQEWL 71
 P G +G AY + NG+++F+K N+++P L L+ E I P+L+W +R NGD+++AQ W+
 Sbjct: 20 PAGGATGDAYYAKH-NGQQLFLKRNSPFLAVLSAEGIVPKLVWTKRMENGDVITAQHWM 78
 Query: 72 DGRITLKEDMGSKQIIHILLRLHKSRLPLVNLQLQGYKIENPYDLLMDWEKQTPIQIREN 131
 GR L +DM + + +L ++H S+ L++ L +LG + NP LL ++ + +
 Sbjct: 79 TGRELKPKDMSGRFVAELLRKIHTSKALLDMLKRLGKEPLNPGALLSQLKQAVFAVQSS 138
 Query: 132 TYLQSIIVTELKRSLEPEFRTEVATIVHGDIKHSNWVITTSGLIYLVWDVSVRLTDRMYDVA 191
 +Q + L+ L E + H D+ H+NW+++ +YL+DWD + D D+
 Sbjct: 139 PLIQEGIKYLEEHLHEVHFGEKVVCHCDVNHNNWLLSEDNQLYLIDWDGAMIADPAMDLG 198
 Query: 192 YILSHYIPQKHWDWLSYYGYKDNEKVWSKIIWY 225
 +L HY+ + W+ WLS YG + E + ++ WY
 Sbjct: 199 PLLYHYVEKPAWESWLSMYGIELTESLRLMAWY 232

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3465> which encodes the amino acid sequence <SEQ ID 3466>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2686(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-1252-

Identities = 214/262 (81%), Positives = 242/262 (91%)

5 Query: 1 MTISNQELTLTPLRGKSGKAYIGTYPNGERVFKYNTTPILPALAKEQIAPQLLWARRTS 60
+T + QELTLTPLRGKSGKAY GTYPNGE VF+K NTTPILPALAKEQIAPQLLWA+R
Sbjct: 1 VTTTEQELTLTPLRGKSGKAYKGTYPNGECVFIKLNTPILPALAKEQIAPQLLWAKRMG 60

10 Query: 61 NGDMMSAQEWLDGRTLTKEDMGSKQIIHILLRLHKSRLVNQLQLGYKIENPYDLLMDW 120
NGDMMSAQEWL+GRTLTKEDM SKQIIHILLRLHKS+ LVNQLQL YKIENPYDLL+D+
Sbjct: 61 NGDMMSAQEWLNGRTLTKEDMNSKQIIHILLRLHKSRLVNQLQLNYKIENPYDLLVDF 120

15 Query: 121 EKQTPIQIRENTYLSIVTELKRSLEPFRTREVATIVHGDIKHSNWVITTSGLIYLVWDWS 180
E+ P+QI++N+YLO+IV ELKRSLEPF++EVATIVHGDIKHSNWVITTS+I+LVDWDS
Sbjct: 121 EQNAPLQIQNSYLAIVKELKRSLEPFKSEVATIVHGDIKHSNWVITTSGMIFLVDWDS 180

20 Query: 181 VRLTDRMYDVAYILSHYIPQKHWDWLSYYGYKDNEKVWSKIWIYGQFSYLSQIIKCFDK 240
VRLTDRMYDVAY+LSHYIP+ W +WLSYYGYK+N+KV KIIWYGQFS+L+QI+KCFDK
Sbjct: 181 VRLTDRMYDVAYLLSHYIPRSRWSEWLSYYGYKNNDKVMQKIWIYGQFSHLTQILKCFDK 240

25 Query: 241 RDMEHVNQEIIYELRKRFRELIKK 262
RDMEHVNQEIIY LRKFRE+ +K
Sbjct: 241 RDMEHVNQEIIYALRKRFREIFRK 262

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1121

A DNA sequence (GBSx1196) was identified in *S.agalactiae* <SEQ ID 3467> which encodes the amino acid sequence <SEQ ID 3468>. Analysis of this protein sequence reveals the following:

30 Possible site: 51
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4529(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00285 GB:AF008220 YtmQ [Bacillus subtilis]
Identities = 126/211 (59%), Positives = 161/211 (75%)

40 Query: 1 MRVRKRKGAEHLENNPHYVISNPEEAKGRWHEIFGNPNPIHIEVSGSGKGAFITGMAEQN 60
MR+R + A++ L N ISNP + KG+W+ +FGN+NPIHIEVG+GKG FI+GMA+QN
Sbjct: 1 MRMRHKPWADDFLAENADIAISNPADYKKGWNTVFGNDNPIHIEVGTGKGQFISGMAKQN 60

45 Query: 61 PDINYIGIDIQLSVLSYALDKVLDGAKNIKLLVDGSSLSNYFDTGEDLMYLNFSDPW 120
PDINYIGI++ SV+ A+ KV DS A+N+KLL +D +L++ F+ GEV +YLNFSDPW
Sbjct: 61 PDINYIGIELFKSVIVTAVQKVKDSEAQNVKLLNIDADTLTDVFEPGEVKRVYLNFSDPW 120

50 Query: 121 PKKHEKRRLTYKTFLDYKDILPEQGEIHFKTDNRGLFEYSLASFQYGMTLKQVWLDL 180
PKK+HEKRRLTY FL Y++++ + G IHFKTDNRGLFEYSL SFS+YG+ L V LDL
Sbjct: 121 PKKRHEKRRLTYSHFLKKYEEVMGKGSIHFKTDNRGLFEYSLKFSFYGLLLTYVSLDL 180

55 Query: 181 HASDYQQNIMTEYERKFSNKGQVIYRVEARF 211
H S+ + NIMTEYE KFS GQ IYR E +
Sbjct: 181 HNSNLEGNIMTEYEEKFSALGQPIYRAEVEW 211

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3469> which encodes the amino acid sequence <SEQ ID 3470>. Analysis of this protein sequence reveals the following:

60 Possible site: 29
>>> Seems to have no N-terminal signal sequence

-1253-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3303(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 179/211 (84%), Positives = 193/211 (90%)

10

Query: 1 MRVRKRKGAEHLNNPHYVISNPEEAKGRWHEIFGNNNPIHIEVGSGKGAFITGMAEQN 60
 MRVRKRKGAEHL NNPHYVI NPE+AKGRWH++FGN+ PIHIEVGSGKG FITGMA +N
 Sbjct: 1 MRVRKRKGAEHLANNPHYVIINPEDAKGRWHDVFGNDRPIHIEVGSGKGGFITGMALKN 60

15

Query: 61 PDINYIGIDIQLSVLSYALDKVLDSGAKNIKLLLDGSSLSNYFDTGEVDLMYLNFSDPW 120
 PDINYIGIDIQLSVLSYALDKVL S N+KLL VDGSSL+NYF+ GEVD+MYLNFSDPW
 Sbjct: 61 PDINYIGIDIQLSVLSYALDKVLASEVVPNVKLLRVDGSSLTNYFEDGEVDMMYLNFSDPW 120

20

Query: 121 PKKKHEKRRLTYKTFLDITYKDILPEQGEIHFKTDRGLFEYSLASFQYGMTLQVWLDL 180
 PK KHEKRRLTYK FLDITYK ILPE GEIHFKTDRGLFEYSLASFQYGMTL+Q+WLDL
 Sbjct: 121 PKTKHEKRRLTYKDFLDITYKRILPEHGEIHFKTDRGLFEYSLASFQYGMTLRQIWLDL 180

Query: 181 HASDYQQNIMTEYERKFSNKGQVIYRVEARF 211
 HAS+Y+ N+MTEYE KFSNKGQVIYRVEA F
 Sbjct: 181 HASNYEGNVMTEYEKFSNKGQVIYRVEANF 211

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1122

A DNA sequence (GBSx1197) was identified in *S.agalactiae* <SEQ ID 3471> which encodes the amino acid sequence <SEQ ID 3472>. Analysis of this protein sequence reveals the following:

30

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35

bacterial cytoplasm --- Certainty=0.1311(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40

>GP:BA06136 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 61/124 (49%), Positives = 81/124 (65%), Gaps = 2/124 (1%)

45

Query: 2 GGDYVLSILIDKPGGITVEDTAQLTDVVSPLLDITQPDFFPEQYMLEVSSPGLERPLKTA 61
 G D+ L + ID G+ +ED ++++ +S LD + DP + Y LEVSSPG ERPLK
 Sbjct: 33 GKDWFLRVFIDSETGVDLEDCKGVSERLSEKLD--ETDPIEQAYFLEVSSPGAERPLKRE 90

Query: 62 EALSNAVGSYINVSLYKSIDKVKIFEGDLLSFDGETLTIDYMDKTRHKTVDIPYQTVAKA 121
 + L ++G ++V+LY+ ID K EG+L FDGETLTI+ KTR KTV IPY VA A
 Sbjct: 91 KDLLRSIGKNVHVTLYEPIDGKALEGELTEFDGETLTIEIKIKTRKKTVTIPYAKVASA 150

50

Query: 122 RLAV 125
 RLAV
 Sbjct: 151 RLAV 154

55

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3473> which encodes the amino acid sequence <SEQ ID 3474>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1254-

bacterial cytoplasm --- Certainty=0.3445(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 101/127 (79%), Positives = 117/127 (91%)

Query: 1 MGGDYVLSILIDKPGGITVEDTAQLTDVVSPLLDTIQDPDFPEQYMLEVSSPGLERPLKT 60
 MG DY+LSIL+DK GGITVEDT+ LT+++SPLLDTI PDPFP QYMLEVSSPGLERPLKT
 10 Sbjct: 52 MGSDYILSILVDKEGGITVEDTSDLTNIISPLLDTIQDPDFNQYMLEVSSPGLERPLKT 111
 Query: 61 AEALSNVAGSYINVS LYKSIDKVKIFEGDLLSFDGETLTIDYMDKTRHKTVDPYQTVAK 120
 A++L AVGSYINVS LY++IDKVK+F+GDLL+FDGETLTIDY+DKTRHK V+IPYQ VAK
 15 Sbjct: 112 ADSLKAAGSYINVS LYQAIDKVKVFGDLLAFDGETLTIDYLDKTRHKIVNIPYQAVAK 171
 Query: 121 ARLAVKL 127
 R+AVKL
 Sbjct: 172 VRMAVKL 178

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1123

A DNA sequence (GBSx1198) was identified in *S.agalactiae* <SEQ ID 3475> which encodes the amino acid sequence <SEQ ID 3476>. This protein is predicted to be a utilization substance protein a homolog (nusA). Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.5069(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9565> which encodes amino acid sequence <SEQ ID 9566> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13533 GB:Z99112 nusA [Bacillus subtilis]
 Identities = 164/370 (44%), Positives = 251/370 (67%), Gaps = 15/370 (4%)
 40 Query: 4 MSKEMLEAFRILEEKHKINKEDIIDAVTESLKSAYKRRYQSESCVIEFNEKKADFTVYT 63
 MS E+L+A ILE+EK I+KE II+A+ +L SAYKR + Q+++ ++ N + V+
 Sbjct: 1 MSSELDAITILEKEKGISKEIIEAIEAALISAYKRNFNQANVRVDLNRGTGSIRVFA 60
 45 Query: 64 VREVVDVDFSRLEISLKDALAISAYELGDKIRFEESVTEFGRVAAQSAKQTIMEKMRR 123
 ++VVDEV+D RLEIS+++A I Y +GD + E + +FGR+AAQ+AKQ + +++R
 Sbjct: 61 RKDVVDEVYDQRLIEISIEEAQGIHPEYMGVDVVEIEVTPKDFGRIAAQTAQVVTQVRVRE 120
 Query: 124 QMREVFNEYKQHEGEIMTGTVERFDORFIYVNLGSLEAQLSHQDQIPGESFKSHDMIDV 183
 R V ++E+ E +IMTG V+R D +FIYV+LG +EA L +Q+P ES+K HD I V
 50 Sbjct: 121 AERGVIYSEFIDREEDIMTGVQLDNKFIYVSLGKIEALLPVNEQMPNESYKPHDRIKV 180
 Query: 184 YVYKVENNPKGVNVSRSHPFIKRIKIMEREIPEVFDGTVEIMSVSREAGDRTKVAVRSH 243
 Y+ KVE KG ++VSR+HP +KR+ E E+PE++DGTVE+ SV+REAGDR+K++VR+
 Sbjct: 181 YITKVEKTTKGPQIYVSRTHEGLLRLFEIEVFIEYDGTVELKSVAREAGDRSKISVRTD 240
 55 Query: 244 NSNVDAIGTIVGRGGSNIKKVISNFHPKRVDAGTGLEIPVEENIDVIQWVEDPAEFITYNA 303
 + +VD +G+ VG G ++ +++ E ID++ W DP EF+ NA
 Sbjct: 241 DPDVDFVSGCVGPKGQVRVQAIVNELK-----GEKIDIVNWSSDPVEFVANA 286

-1255-

Query: 304 IAPAEVDMVLFDDDEDTKRATVVVPDSKLSLAIGRRGQNVRLAAHLTG YRIDIKSASEYEK 363
 ++P++V V+ ++E+ K TV+VPD +LSLAIG+RGQN RLAA LTG++IDIKS ++ +
 Sbjct: 287 LSPSKVLDVIVNEEE-KATTIVVPDYQLSLAIGKRGQNARLAAKLTGWKIDIKSETDARE 345

Query: 364 MEAQELQTEE 373
 + + EE
 Sbjct: 346 LGIYPRELEE 355

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3477> which encodes the amino acid sequence <SEQ ID 3478>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2074 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 348/380 (91%), Positives = 361/380 (94%), Gaps = 2/380 (0%)

- Query: 4 MSKEMLEAFRILEEKHINKEDIIDAVTESLKSAYKRRYGQSESCVIEFNEKKADFTVYT 63
 MSKEMLEAFRILEEKKHI+K DIIDAVTESLKSAYKRRYGQSESCVIEFNEK ADF V+T
- 25 Sbjct: 12 MSKEMLEAFRILEEKHIDKADIIDAVTESLKSAYKRRYGQSESCVIEFNEKTADFQVFT 71
- Query: 64 VREVVEVFDVSRLEISLKDALAISSAYELGDKIRFEESVTEFGRVAAQSAKQTIMEKMRR 123
 VREVVEVFDVSRLEISLKDALAISSAYELGDKIRFEESV EFGVAAQSAKQTIMEKMRR
- 30 Sbjct: 72 VREVVEVFDVSRLEISLKDALAISSAYELGDKIRFEESVNEFGRVAAQSAKQTIMEKMRR 131
- Query: 124 QMREVTFNQYKQHEGEIMTGTVERFDQRFIYNLGSLEAQLSHQDQIPGESFKSHDMIDV 183
 QMREV FNEYK+HEGEIMTGTVERFDQRFIYNLGSLEAQLSHQDQIPGE+FKSHD IDV
- 35 Sbjct: 132 QMREVMFNEYKEHEGEIMTGTVERFDQRFIYNLGSLEAQLSHQDQIPGETFKSHDRIDV 191
- Query: 184 VYVKVENNPKG VNVFVSRSHPEFIKRIMEE+EIPEVFDGTVEIMSVSREAGDRTKVAVRSH 243
 VYVKVENNPKG VNVFVSRSHPEFIKRIME+EIPEVFDGTVEIMSVSREAGDRTKVAVRSH
- 40 Sbjct: 192 VYVKVENNPKG VNVFVSRSHPEFIKRIMEQEIPEVFDGTVEIMSVSREAGDRTKVAVRSH 251
- Query: 244 NSNVDAGTIVGRGGSNIKKVISNFHPKRVDAKTGLEIPVEENIDVIQWVEDPAEFIYNA 303
 N NVDAIGTIVGRGGSNIKKVIS FHPKRVDAKTGLEIPVEENIDVIQWV+DPAEFIYNA
- 45 Sbjct: 252 NPNVDAGTIVGRGGSNIKKVISKFHPKRVDAKTGLEIPVEENIDVIQWVDDPAEFIYNA 311
- Query: 304 IAPAEVDMVLFDDDEDTKRATVVVPDSKLSLAIGRRGQNVRLAAHLTG YRIDIKSASEYEK 363
 IAPAEVDMVLFDDDEDT KRATVVVPDSKLSLAIGRRGQNVRLAAHLTG YRIDIKSASEY++
- 50 Sbjct: 312 IAPAEVDMVLFDDDEDLKRATVVVPDSKLSLAIGRRGQNVRLAAHLTG YRIDIKSASEYDR 371
- Query: 364 MEAQELQTEEVAQESEVISD 383
 +EA+ + A E V+ D
- Sbjct: 372 LEAE--KEAATAVEEPVVD 389

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1124

- 55 A DNA sequence (GBSx1199) was identified in *S.agalactiae* <SEQ ID 3479> which encodes the amino acid sequence <SEQ ID 3480>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

- 60 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2012 (Affirmative) < succ>

-1256-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB13534 GB:Z99112 alternate gene name: ymxB~similar to
 hypothetical proteins [Bacillus subtilis]
 Identities = 46/92 (50%), Positives = 67/92 (72%), Gaps = 1/92 (1%)
 10 Query: 1 MAKTKKIPLRKSVVSGEVIDKRDLLRIVKNKEGQVFIDPTGKQNGRGAYIKLDNDEAILA 60
 M K KKIPLRK VV+GE+ K++L+R+V++KEG++ +DPTGK+NDRGAY+ LD + + A
 Sbjct: 1 MNKHKKIPLRKCVVTGEMKPKKELIRVVRSEKEGISVDPTGKKNRGAYLTLDKECILAA 60
 Query: 61 KKKRVFDRSFSMEVSDEFYDELLAYVDHKVKR 92
 KKK F ++ D+ +DELL + KVK+
 15 Sbjct: 61 KKKNTLQNFQSQIDDDQIFDELLELAE-KVKK 91

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3481> which encodes the amino acid sequence <SEQ ID 3482>. Analysis of this protein sequence reveals the following:

20 Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1008 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 77/98 (78%), Positives = 92/98 (93%)
 30 Query: 1 MAKTKKIPLRKSVVSGEVIDKRDLLRIVKNKEGQVFIDPTGKQNGRGAYIKLDNDEAILA 60
 M+K KKIPLRK+VSGE+I KRDLLRIVK K+GQVFIDPTGKQNGRGAYIKLDN EA++A
 Sbjct: 2 MSKVKKIPLRKSLVSGEIIAKRDLLRIVKTKDQGVFIDPTGKQNGRGAYIKLDNQEALMA 61
 Query: 61 KKKRVFDRSFSMEVSDEFYDELLAYVDHKVKRRELGLE 98
 KKK+VF+RSFSM++ + FYD+L+AYVDHK+KRRELGL+
 35 Sbjct: 62 KKKQVFNRFSMDIPESFYDDLIAVVDHKIKRRELGLD 99

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1125

A DNA sequence (GBSx1200) was identified in *S.agalactiae* <SEQ ID 3483> which encodes the amino acid sequence <SEQ ID 3484>. This protein is predicted to be probable ribosomal protein in infb 5' region. Analysis of this protein sequence reveals the following:

45 Possible site: 19
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:BAB06133 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 46/95 (48%), Positives = 65/95 (68%), Gaps = 1/95 (1%)
 Query: 6 KVLNLIGLAQRAGRLITGEELVIKAIQNQQVSLIFLANDAGPNLTKKVTDKSNYYKTEVS 65
 K L+L+GLA RA +L+TGEE V+KA+QN QV+L+ L++DAG + KK+ DK Y+ V

-1257-

Sbjct: 5 KWLSSLGLAARARQLLTGEEQVVKAVQNGQVTLVILSSDAGIHTKKLLDKCGSYQIPVK 64

Query: 66 TVFSTLELSDALGK-PRKVVAVADAGFSKMMRTLM 99

V + L A+GK R V+ V DAGFS+K+ L+

5 Sbjct: 65 VVGNRQMLGRAIGKHERVVIGVKDAGFSRKLALTI 99

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3485> which encodes the amino acid sequence <SEQ ID 3486>. Analysis of this protein sequence reveals the following:

Possible site: 45

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1950(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 75/99 (75%), Positives = 88/99 (88%)

20 Query: 1 MNNSEKVLNLIGLAQRAGRLITGEELVVIKAIQNQQVSLIFLANDAGPNLTKKVTDKSNYY 60
+ N E++ +LIG AQRAG++I+GEELV+KAIQ+QQV L+FLANDAGPN+TKKVTDKSNYY

Sbjct: 1 LTNLERLSSLIGPAQRAGKVISGEELVVKAIQHQQVILVFLANDAGPNVTKKVTDKSNYY 60

Query: 61 KTEVSTVFSTLELSDALGKPRKVVAVADAGFSKMMRTLM 99

25 EVSTV + LELS ALGKPRKV A+ADAGFSKMMRTLM

Sbjct: 61 NTEVSTVLNLELSAALGKPRKVAIADAGFSKMMRTLM 99

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 1126

A DNA sequence (GBSx1201) was identified in *S.agalactiae* <SEQ ID 3487> which encodes the amino acid sequence <SEQ ID 3488>. Analysis of this protein sequence reveals the following:

Possible site: 37

35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2873(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10959> which encodes amino acid sequence <SEQ ID 10960> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3489> which encodes the amino acid sequence <SEQ ID 3490>. Analysis of this protein sequence reveals the following:

45 Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2985(Affirmative) < succ>

50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 735/961 (76%), Positives = 805/961 (83%), Gaps = 42/961 (4%)

55

-1258-

Query: 1 MSKRLHEIAKEIGKTSKEVVEQAQSLGLPVKSHASSVEENDATRIVESFS-SSKTKAPT 59
 +SKKRLHEIAKEIGK+SKEVVE A+ LGL VKSHASSVEE DA +I+ SFS +SK
 Sbjct: 1 LSKRLHEIAKEIGKSSKEVVEHAKYLGLDVKSHASSVEEADAKI ISSFSKASKPDVTA 60

5 Query: 60 NSVQTNQGVKTESKTVETKQGLSDDKPSTQPVAKPKPQSRNFKAEREARAKAEAEKRQHN 119
 + + V S TV + G S+ TQ V+KPK SRNFKAEREARAK +A ++Q N
 Sbjct: 61 SQTVPKPEVAQPSVTTVVKETG-SEHVEKTQ-VSKPK--SRNFKAEREARAKEQAARKQAN 116

10 Query: 120 GD-----HRKNNRHNDRSDDRR--HQGQKRSNGNR-----NDNRQ--G 154
 G +R+ N H D+R H+ Q +N R +DN Q G
 Sbjct: 117 GSSHRSQERRGGYRQPNNHQTNQEGDKRITHRSQGDNDKRIERKASNVSPRHDHNLQVVG 176

15 Query: 155 QQNN---RNKNDGRYADHKQKPQTRPQQAGNRIDFKARAAALKAEQNAEYSRHSEQR 210
 +N N +GR+ + K++ + PQ + +IDFKARAAALKAEQNAEYSR SE RF
 Sbjct: 177 DRNRSAFENHKNRFTNQKKQGRQEPQSKSP-KIDFKARAAALKAEQNAEYSRQSETRF 235

20 Query: 211 REEQEAKRQAAKEQELAKAAALKAEQAEQAKEKLASKPVAKVKEIVNKVAATPSQTADS 270
 R +QEAKR A ++ AK AALKAQ E +E A K + + + + TAD+
 Sbjct: 236 RAOQEAKRLAELARQEAKEALKAQAEEMSHREA-ALKSIEEAETKLKSSNISAKSTADN 294

25 Query: 271 RRRKQTRSDKSRQFSNENEDGQKQTRNKKNWNNQNVNRQNSNWNHKNKKGK----T 326
 RRRKQ R +K+R+ ++ ++GQK +NKK+WN+QNQVRNQ+NSNWN NKK KKGK T
 Sbjct: 295 RRRKQARPEKNRBLTHHSQEGQK--KNKKSWSNQNVNRQKNSNWNKKNKTKKGKKNVKN 352

30 Query: 327 NGAPKPVTERKFHELPEFEYTEGMTVAEIAKRIKREP AEIVKKLFMMGVMATQNQSLDG 386
 N APKPVTERKFHELPEFEYTEGMTVAEIAKRIKREP AEIVKKLFMMGVMATQNQSLDG
 Sbjct: 353 NTAPKPVTERKFHELPEFEYTEGMTVAEIAKRIKREP AEIVKKLFMMGVMATQNQSLDG 412

35 Query: 387 DTIELLMVDYGIEAHAKVEVDEADIERFFADEYLNPDNLTERPPVVTIMGHVDHGKTTL 446
 DTIELLMVDYGIEA AKVEVD+ADIERFF DE+YLNPN+N+ ER PVVTIMGHVDHGKTTL
 Sbjct: 413 DTIELLMVDYGIEAKAKVEVDDADIERFFEDENYLNPNIVERAPVVTIMGHVDHGKTTL 472

40 Query: 447 LDTLRNSRVATGEAGGITQHIGAYQIEEAGKKITFLDTPGHAAFTSMRARGASVTDITIL 506
 LDTLRNSRVATGEAGGITQHIGAYQIEEAGKKITFLDTPGHAAFTSMRARGASVTDITIL
 Sbjct: 473 LDTLRNSRVATGEAGGITQHIGAYQIEEAGKKITFLDTPGHAAFTSMRARGASVTDITIL 532

45 Query: 507 IVAADDGVMPQTVEAINHNSKAAGVPIIVAINKIDKPGANPERVISELAEHGVISTAWGGE 566
 IVAADDGVMPQT+EAINHNSKAAGVPIIVAINKIDKPGANPERVI+ELAE+G+ISTAWGGE
 Sbjct: 533 IVAADDGVMPQTIEAINHNSKAAGVPIIVAINKIDKPGANPERVIAELAEYGIISTAWGGE 592

50 Query: 567 SEFVEISAKFGKNIQELLETVLLVAEMEELKADADVRAIGTVIEARLDKKGAVATLLVQ 626
 EFVEISAKF KNI ELLETVLLVAE+EELKAD VRAIGTVIEARLDKKGGA+ATLLVQ
 Sbjct: 593 CEFVEISAKFNKNIDELLETVLLVAEVEELKADPTVRAIGTVIEARLDKKGGAATLLVQ 652

55 Query: 627 QGTLNVQDPIVVGNTFGRVRAMTNDLGRVRVKVAGPSTPVSITGLNEAPMAGDHFAYVADE 686
 QGTL+VQDPIVVGNTFGRVRAM NDLGRVRK A PSTPVSITGLNE PMAGDHFAYVADE
 Sbjct: 653 QGTLHVQDPIVVGNTFGRVRAMVNDLGRVRKSAEPSTPVSITGLNETPMAGDHFAYVADE 712

60 Query: 687 KAARAAGEERAKRALLKQRQNTQVRSLNLFDTLKAGEVKSNNVIKADVQGSVEALAAS 746
 KAARAAGEER+KRALLKQRQNTQVRSL+NLFDTLKAGE+K+VNNVIKADVQGSVEALAAS
 Sbjct: 713 KAARAAGEERSKRALLKQRQNTQVRSLNLFDTLKAGEIKTVNNVIKADVQGSVEALAAS 772

65 Query: 747 LLKIDVEGVKVN NVHSAVGAINESDVTLEASNAVIIGFNVVRPTPQARQQADDDVEIR 806
 L+KI+VEGV+VNVHSAVGAINESDVTLEASNAVIIGFNVVRPTPQARQQAD DDVEIR
 Sbjct: 773 LVKIEVEGVVRNV NVHSAVGAINESDVTLEASNAVIIGFNVVRPTPQARQQADDDVEIR 832

Query: 807 HSIIYKVIEEVEEAMKGKLDPEYQEKILGEAIIRETFKVSKVGTIGGFMVINGKVTRDSS 866
 HSIIYKVIEEVEEAMKGKLDP YQEKILGEAIIRETFKVSKVGTIGGFMVINGKVTRDSS
 Sbjct: 833 HSIIYKVIEEVEEAMKGKLDPVYQEKILGEAIIRETFKVSKVGTIGGFMVINGKVTRDSS 892

Query: 867 VRVIRGCVVIFDGKSLASLKHKKDDVKEVGNAQEGGLMIENYNDLKEDDTIEAYIMEEIKRK 927
 VRVIRG VVIFDGKSLASLKHKKDDVKEVGNAQEGGLMIEN+NDLK DDTIEAYIMEEIRK
 Sbjct: 893 VRVIRDSVVIFDGKSLASLKHKKDDVKEVGNAQEGGLMIENFNDLKVDVDTIEAYIMEEIVRK 953

65 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1127

A DNA sequence (GBSx1202) was identified in *S.agalactiae* <SEQ ID 3491> which encodes the amino acid sequence <SEQ ID 3492>. This protein is predicted to be ribosome binding factor A (rbfA). Analysis of this protein sequence reveals the following:

5 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2557(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9567> which encodes amino acid sequence <SEQ ID 9568> was also identified.

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3493> which encodes the amino acid sequence <SEQ ID 3494>. Analysis of this protein sequence reveals the following:

 Possible site: 60
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.4765(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below.

 Identities = 93/117 (79%), Positives = 103/117 (87%)

 Query: 8 LIMANHRIDRVGMEIKREVNEILRLRVNDPRVQDVTTITDVQMLGDLMAKVFTIHTSLA 67
 + MANHRIDRVGMEIKREVN+IL+ +V DPRVQ VTIT+VQM GDL+AKV+YTI S LA
 30 Sbjct: 1 MAMANHRIDRVGMEIKREVNDILQKKVRDPRVQGVTTITEVQMGGDL+LAKVYYTIMSDLA 60

 Query: 68 SDNQKAQIGLEKATGTIKRELKGNLTMYKIPDLQFVKDESIEYGNKIDEMLRNLDKK 124
 SDNQKAQ GLEKATGTIKRELK LTMYPKIPDL F KD SI YGNKID++LR+LD K
 35 Sbjct: 61 SDNQKAQTGLEKATGTIKRELKQLTMYKIPDLVFEKDNSIAYGNKIDQLLRDLDNK 117

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1128

40 A DNA sequence (GBSx1203) was identified in *S.agalactiae* <SEQ ID 3495> which encodes the amino acid sequence <SEQ ID 3496>. This protein is predicted to be esterase. Analysis of this protein sequence reveals the following:

 Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.

45 ----- Final Results -----

 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database.

 >GP:AAA79277 GB:M64783 acetyl-hydrolase [Streptomyces hygroscopicus]
 Identities = 58/220 (26%), Positives = 90/220 (40%), Gaps = 8/220 (3%)

-1260-

Query: 98 WNDNGKANQKTIIFYLAGGSYLNNPTYPYHISMLKTLSTSLDAKIILPIYPKTPRYTYDYAI 157
 W + + +T+ YL GGSY H + L + A ++ Y + P + A+
 Sbjct: 58 WVRPARQDGRLLYLGGSYALGSEFQSHRHLSSALGDAAGAAVLALHYRRPPESPFPAAV 117

5 Query: 158 PRLVNLRYHFHEKN---ANLTLMGDSAGGGLALGLAHALSHQSGQEAIQPKNIILLSPW 214
 V YR E+ +TL GDSAG GLA+ AL P P + +SPW
 Sbjct: 118 EDAVAAYRMLLEQGCPPGRVTLAGDSAGAGLAVAALQALR---DAGTFLPAAAVCISPW 173

10 Query: 215 LDVTMKHPEIPKYEDTDPILSAWGLARVGEIWANGSNNTNHTYVSPKNAPATKLAPITLF 274
 D+ + + +L L R+ E + G+ + H SP + T L P+ +
 Sbjct: 174 ADLACEGASHTTRKAREILLDTADLRMAERYLAGT-DPRHPLASEAHGDLTGLPPLLIQ 232

Query: 275 TGTREIFFPDIRDYAAQLQAANHPVNYIAQEGMNHVYPIY 314
 G+ E+ D R A PV + M HV+ Y
 15 Sbjct: 233 VGSEEVLDHDDARALEQAALKAGTPVTFEEWPEMFHVWHWY 272

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3497> which encodes the amino acid sequence <SEQ ID 3498>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 244/334 (73%), Positives = 280/334 (83%), Gaps = 6/334 (1%)

30 Query: 1 MKPSFKKLLLLFSIITILSIACTPHAKASGRSWKSWFIEQYFWLKRDKSYKQDESSFQ 60
 +K +K L+ ++ L + TP A AS RSWKSWFIEQYFWLKRDKSYY QD+ SFQ
 Sbjct: 1 LKHPIRKTLLVTLGLLLTLCLP-TPVA-ASSRSWKSWFIEQYFWLKRDKSYYSKQDDPSFQ 58

Query: 61 KYLNASREQSDKGYLDPNVNGGLVQERLFDQVYSWNDNGKANQKTIIFYLAGGSYLNN 120
 +YL+A REQSDK Y LD N VNG LVQE L+ MQVYSWNDNGK +QKTI YLAGGSYLNN
 35 Sbjct: 59 RYLDACREQSDKPYQLDTNLVNGPLVQENLYGMQVYSWNDNGKPDQKTIIFYLAGGSYLNN 118

Query: 121 PTPYHISMLKTLSTSLDAKIILPIYPKTPRYTYDYAIPRLVNLRYHFHEKNANLTLMGDS 180
 PT YHI+MLKTLSTSLDAKI+LPIYPK PRYTY+Y +P+LVNLY+H++ KN N+ LMGDS
 40 Sbjct: 119 PTTYHINMLKTLSTSLDAKIVLPIYPKAPRYTYNYTMPKLVNLYQHYYHKNQNVFLMGDS 178

Query: 181 AGGGLALGLAHALSHQSGQEAIQPKNIILLSPWLDVTMKHPEIPKYEDTDPILSAWGLA 240
 AGGGLALGLAHAL + E++PQPK ++LLSPWLDVTM HPEIP+YED DPILS+WGL
 Sbjct: 179 AGGGLALGLAHALHN---ESVPQPKQLVLLSPWLDVTMSHPEIPEYEDADPILSSWGLK 234

45 Query: 241 RVGEIWANGSNNTNHTYVSPKNAPATKLAPITLFTGTREIFFPDIRDYAAQLQAANHPVN 300
 RVGE+WA ++NTNH YVSPKN P T L PITLFTGTREIF+PDIRDYAA+L+AAANH +
 Sbjct: 235 RVGELWAYSADNTNHTYVSPKNGPITYLPPITLFTGTREIFYPDIRDYAAKLKAANHNIT 294

50 Query: 301 YIAQEGMNHVYPIYPIEEAKTAQYQMIDIINKTP 334
 +I QEGMNHVYPIYPIEEAKTAQYQ+ID INKTP
 Sbjct: 295 FITQEGMNHVYPIYPIEEAKTAQYQIIDAINKTP 328

A related GBS gene <SEQ ID 8731> and protein <SEQ ID 8732> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6
 McG: Discrim Score: 11.88
 GvH: Signal Score (-7.5): -1.33
 Possible site: 28
 60 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 4.03 threshold: 0.0
 PERIPHERAL Likelihood = 4.03 174
 modified ALOM score: -1.31

-1261-

*** Reasoning Step: 3

----- Final Results -----

```
5      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

10 28.4/46.2% over 220aa

Streptomyces

hygroscopicus

EGAD	5925	acetyl-hydrolase Insert characterized
------	------	---------------------------------------

```

15      ORF00486(589 - 1245 of 1602)
      EGAD|5925|5724(57 - 277 of 300) acetyl-hydrolase {Streptomyces hygroscopicus}
      %Match = 6.8
      %Identity = 28.3  %Similarity = 46.1
      Matches = 62  Mismatches = 111  Conservative Sub.s = 39

```

[illegible]

35

927	957	987	1017	1047	1077	1107	1137
EAIPQPKNIILLSPWLDVTMKHPEIPKYEDTDPILSAWGLARVGEIANGSNNTNHTYVSPKNAPATKLAPITLFTGTRE							
: : : : : : : : : : : : :							
AGTPLPAAAVCISFWADLACEGASHTTRKAREILLDTADLRMEAERYLAGTD-PRHPLASPAHGDLTGLPPLLIQVGSSE							
170	180	190	200	210	220	230	

40

1167	1197	1227	1245	1275	1305	1335	1365
IFFPDIRDYAAQLQAANHVPVNYIAQEGMNHV---YPIYPIEBEAKTAQYQMIDIINKTP*Y*LSQL*SYKK*TMILTWF							
:	:	:	:	:	:	:	:
VLHDDARALEQAALKAGTPTVTFEEWPEMFHVHWHYHPVLPEGRRAAIEVAGAFRLRTATGEGGLK							
250	260	270	280	290	300		

45 SEQ ID 8732 (GBS149) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 6; MW 37kDa).

The GBS149-His fusion product was purified (Figure 196, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 291), which confirmed that the protein is immunoaccessible on GBS bacteria.

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1129

55 A DNA sequence (GBSx1204) was identified in *S.agalactiae* <SEQ ID 3499> which encodes the amino acid sequence <SEQ ID 3500>. This protein is predicted to be CopY. Analysis of this protein sequence reveals the following:

Possible site: 22

```
>>> Seems to have no N-terminal signal sequence
```

-1262-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3140(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAG10085 GB:AF296446 CopY [Streptococcus mutans]
 Identities = 67/137 (48%), Positives = 98/137 (70%)

Query: 2 TISSAEWEIMRVVWAQQNTTSNEILAVLLEKYDWTPTVKTLLRRLLDKGYVSREKMGKG 61
 +IS+AEWE+MRVVWA+Q T+S+EI+A+L Y W+ ST+KTL+ RL +KGY++ ++ G+
 Sbjct: 3 SISNAEWEVMRVVWAKQMTSSSEIIAILSRTYCWSASTIKTLITRLSEKGYLTSQRQGRK 62

15 Query: 62 FSYSPLIDEDLMMSEVDSVFQKVCQTKHVAIVRHLESIPMTEKDRNLQSSLEAKKGK 121
 + YS LI E+ A+ +V VF ++C TKH A++RHL+E PMT D L++ L +KK
 Sbjct: 63 YIYSSLISEEEALEQQVSEVFSRICVTKHQALIRHLVEETPMTLSDIEKLEALLSKKAN 122

20 Query: 122 TLERVACNCIPGQCQC 138
 + V CNCI GQC C+
 Sbjct: 123 AVPEVKCNCIVGQCSCY 139

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3501> which encodes the amino acid sequence <SEQ ID 3502>. Analysis of this protein sequence reveals the following:

25 Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2331(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 54/135 (40%), Positives = 84/135 (62%)

Query: 3 ISSAEWEIMRVVWAQQNTTSNEILAVLLEKYDWTPTVKTLLRRLLDKGYVSREKMGKGF 62
 IS+AEWE+MRVVWA + S++I+ +L +KY W+ ST+KTL+ RL+ K +++ + G+ +
 Sbjct: 10 ISAEWEVMRVVWASGDIKSSDIITILRKKYQWSDSTIKTLIGRLVKKNFLTYSYRQGRAY 69

40 Query: 63 SYSPLIDEDLMMSEVDSVFQKVCQTKHVAIVRHLESIPMTEKDRNLQSSLEAKKGKT 122
 Y L+DE L + +V +CQ +H ++ L +PMT ++ Q LE KK
 Sbjct: 70 IYQALLDETLLQKEALATVLDGICQQRHTRLLRLYHLPMTEIGAFQELLEVKKENA 129

45 Query: 123 LERVACNCIPGQCQC 137
 + V CNC+PGQC C
 Sbjct: 130 VLEVPCNCLPGQCHC 144

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 **Example 1130**

A DNA sequence (GBSx1206) was identified in *S.agalactiae* <SEQ ID 3503> which encodes the amino acid sequence <SEQ ID 3504>. This protein is predicted to be CopA. Analysis of this protein sequence reveals the following:

55 Possible site: 19
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.82	Transmembrane	382 - 398 (370 - 406)
INTEGRAL	Likelihood = -8.01	Transmembrane	356 - 372 (344 - 374)
INTEGRAL	Likelihood = -2.50	Transmembrane	719 - 735 (719 - 738)

-1263-

INTEGRAL	Likelihood = -2.28	Transmembrane	202 - 218 (202 - 218)
INTEGRAL	Likelihood = -1.59	Transmembrane	693 - 709 (691 - 712)
INTEGRAL	Likelihood = -1.33	Transmembrane	167 - 183 (167 - 183)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG10086 GB:AF296446 CopA [Streptococcus mutans]
 Identities = 440/740 (59%), Positives = 571/740 (76%), Gaps = 1/740 (0%)

15 Query: 5 KETFLIDGMTASCALTIEKAVNKLDHVD SAVVN LATEKMTVTFDDTTLS PNVIEECVSE 64
 +E FLIDGMTASCASCA+ +E AV KLD ++SAVVNL TEKMT+ +D +S + + V+
 Sbjct: 3 BEVFLIDGMTASCASCAINVENAVKKLDGIESAVVNLTTEKMTIDYDAKVSEADVTTKAVAG 62

20 Query: 65 SGYEASLFKEETSKSQSERHQLAIEKMWHRFWMSAVATIPLLYISMGP MINLWLP SFLMP 124
 +GY A ++ T++SQ +R + + + R +++ TIPL YI+MG M+ L LP+FL P
 Sbjct: 63 AGYGAKVYDPTTAESQKDR EEHLKAGIKKRL LWT SIFTIPLFYI AMGSMVGLPLPNFLAP 122

25 Query: 125 DKGPLNYGMIQLLLTLPVMYFGRIFYQNGFKALFKRHPNMDSLVAIATTA AFYISLYGLY 184
 PL Y M+ LLLT+PV+ FY NGF++L FK HPNMDSLV++ATTAAF+YSLYG Y
 Sbjct: 123 SSAPLTYAMVLLLLTIPVIVLSWSFYDNGFRSLFKGHPNMDSLVSLATTAAFLYSLYGTY 182

30 Query: 185 EILQGDIIHYAHQLYFESVAVILTLITLGKYFEILSKGRTSASIEKLLT LSAKEARVIKDG 244
 + G H+AH LY+ESVAVILTLITLGKYFE LSKGRTS +I+KL+ LSAKEA +I+DG
 Sbjct: 183 HVYLGHTTHAHHLHYESVAVILTLITLGKYFETLSKGRTSDAIKKLMHLSAKEATLIRDG 242

35 Query: 245 EDYMVPLDKVKIGETILVKGPEKIPLDGHV VAGESSIDESMLTGESIPVEKKVGSKVYGA 304
 E+ VP+++V+I + ILVKGPEKIP+DG V++G S+IDESMLTGESIP+EK S VY
 Sbjct: 243 EEIKVPPIEQVQIRDQILVKGPEKIPVDGRVLSGHS AIDESMLTGESIPTEKMDSPVYAG 302

40 Query: 305 SINGQGSLTIFVEKEAGGSLLSQIINLVEAAQTSKAPIANLADK VSGVFVPFVIVIAILS 364
 SINGQGSLT EK +LLSQII LVE AQ +KAPIA +ADKVS VFVP +I IAIL+
 Sbjct: 303 SINGQGSLTFEAEKVGNETTLLSQIIKLVENAQQTKAPIAKIADKVS AVFVPVITITAILT 362

45 Query: 365 GLSWYLILQGSFAFSLKIMIAVLVIACPCALGLATPTAIMVASGKAAENGILFKGGEVLE 424
 GL WY ++CQ F FS+ I +AVLVIACPCALGLATPTAIMV +G+AAENGIL+K G+VLE
 Sbjct: 363 GLFWYFVMGQDFTFSMTISAVLVIACPCALGLATPTAIMVGTGRAAENGILYKRGDVLE 422

50 Query: 425 KAHHIDTIVFDKTGTLTKGKPEVVAIKTYGGDK EEFLGQVASVEKLSNHPLSQTIVNKAK 484
 AH I+TIVFDKTGT+T+GKPEVV +Y D+ + + A++E LS HPLSQ IV+ AK
 Sbjct: 423 LAHQINTIVFDKTGTTTQKGPEVVHQFSY-HDRTDLVQVTAAL EALSEHPLSQ AIVDYAK 481

55 Query: 485 EKELPLREVMAFKNILGYGLSATINGKTM LVGNANLMTKNDVNLDLAKADIEIAQEEAQT 544
 ++ L V F ++ G GL + +T+LVGN LM + +++L+ A+AD + A + QT
 Sbjct: 482 KEGTHLLAVDDFTSLTGLGLKGCVADETLLVGN EKL MRQANISLEQAQADFKAATAQQT 541

60 Query: 545 VVYVSENGVLSGLITLTDQLKTD SQETVKQLQRLG FNLVLLTGDNKASADAIAQKLGITT 604
 ++V+ +G L GLIT+ D++K DS TVK LQ +G + +LTGDN+ +A AIA+++GIT
 Sbjct: 542 PIFVASDQQLGLITLTIADKVKND SAATVKALQNMGVEVAMLTGDNEETAQAIAKEVGITF 601

65 Query: 605 VVSEVLDPQKANVILELKEKGGQIAMVGDGINDAPALASSDVGISMSSGTDIAIESADIV 664
 V+S+V +K IL+L+ +G ++AMVGDGINDAPALA++D+GISM SGT DIA+ESADIV
 Sbjct: 602 VISQVFSQEKQTQAILDLQAEGKKVAMVGDGINDAPALATADIGISMSSGTDIAMESADIV 661

Query: 665 LMKPELTDLLKAMTISKQTIQI IKENLFWAFFYNVLAIPVAMGV LHLFGGP LLNPMLAGL 724
 LMKP + D++KA+ IS+ TI IKENLFWAF YNVL++P+AMGV L+LFGGP LL+PM+AGL
 Sbjct: 662 LMKPAMLDIIKALKISRVTIINIKENLFWAFIYNVLSVP IAMGVLYLFGGP LLDPMIAGL 721

Query: 725 AMAFSSSVSVVLNLR LKVLK 744
 AM+FGSSSVSVVLNLR LKVK+K
 Sbjct: 722 AMSFSSSVSVVLNLR LKVKV 741

There is also homology to SEQ ID 3506.

-1264-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1131

A DNA sequence (GBSx1207) was identified in *S.agalactiae* <SEQ ID 3507> which encodes the amino acid sequence <SEQ ID 3508>. This protein is predicted to be cation-transporting ATPase, P-type (pacS). Analysis of this protein sequence reveals the following:

Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1934 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG10087 GB:AF296446 CopZ [Streptococcus mutans]
Identities = 31/67 (46%), Positives = 43/67 (63%)

Query: 1 MKHTYRVSGMKCDGCAKTVSDKLSSVIGVDEVNVDLTKNQVVVSGKTFKWLKRLKDTK 60
M+ TY + G+KC GCA V+ + S + V++V VDL K +V ++G KW LKR+LK T
Sbjct: 1 MEKTYHIDGLKCKQCADNVTKRFSELKKVNDVKVDLDKKEVRITGNPSKWSLKRALKGTN 60

Query: 61 YSLEEEI 67
Y L EI
Sbjct: 61 YELGAEI 67

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3509> which encodes the amino acid sequence <SEQ ID 3510>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.2997 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 33/63 (52%), Positives = 48/63 (75%)
Query: 1 MKHTYRVSGMKCDGCAKTVSDKLSSVIGVDEVNVDLTKNQVVVSGKTFKWLKRLKDTK 60
M+ Y+V+GM CDGCA+TV++KLS+V GV V V+L K + V+G+ +L+KR+LKDTK
Sbjct: 1 MEKHYQVTGMTCDGCARTVTEKLSAVPGVQSVQVNLEKGEAKVTGRPLTFLIKRALKDTK 60
Query: 61 YSL 63
+ L
Sbjct: 61 FEL 63

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1132

A DNA sequence (GBSx1208) was identified in *S.agalactiae* <SEQ ID 3511> which encodes the amino acid sequence <SEQ ID 3512>. Analysis of this protein sequence reveals the following:

Possible site: 20

-1265-

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -7.59 Transmembrane 67 - 83 (65 - 90)
 INTEGRAL Likelihood = -3.72 Transmembrane 35 - 51 (31 - 51)
 INTEGRAL Likelihood = -3.61 Transmembrane 122 - 138 (120 - 139)
 INTEGRAL Likelihood = -1.59 Transmembrane 154 - 170 (154 - 171)

----- Final Results -----

bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8733> which encodes amino acid sequence <SEQ ID 8734> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 5

McG: Discrim Score: 4.09

GvH: Signal Score (-7.5): 3.87

Possible site: 20

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 4 value: -7.59 threshold: 0.0

INTEGRAL Likelihood = -7.59 Transmembrane 65 - 81 (63 - 88)
 INTEGRAL Likelihood = -3.72 Transmembrane 33 - 49 (29 - 49)
 INTEGRAL Likelihood = -3.61 Transmembrane 120 - 136 (118 - 137)
 INTEGRAL Likelihood = -1.59 Transmembrane 152 - 168 (152 - 169)
 PERIPHERAL Likelihood = 0.85 96

modified ALOM score: 2.02

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15351 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 107/192 (55%), Positives = 137/192 (70%)

Query: 8 WNILSLVGTVAFASSGAIVAIEEEFDILGLFILGFVTAFGGGAIRNVLIGLPIETLWSQG 67
 W +LS++G +AFA SGAIVA+EEE+DILG++ILG VTAFGGGAIRN+LIG+P+ LW QG

Sbjct: 3 WELLSVIGIIAFAVSGAIVAMEEEYDILGVYILGIVTAFGGGAIRNLLIGVPVSALWEQG 62

Query: 68 IAFYAAAAAILFIMIFPNLLSGKGRDAEVVSDAIGLAAFSVQGALYATQSHQPLSAVIVA 127
 F A +I + +FP LL +SDAIGLAAF++QGALYA + PLSAVIVA

Sbjct: 63 AYFQIALLSITIVFLFPKLLKHNKWNKGNLSDAIGLAFAIQGALYAVKMGHPLSAVIVA 122

Query: 128 AVLTGAGGGIVRDVLAGRKPGVLRSEIYAGWSILVGIILYFKIAKTTTDDYLLVLVVTSL 187
 AVLTG+GGGI+RD+LAGRKP VL++EIYA W+ L G+I+ + Y+L V+

Sbjct: 123 AVLTGSGGGIIRDLLAGRKPLVLKARIYAVWAALGGLIVGLGWLGNSTGLYVLFVFLVVC 182

Query: 188 RMLGYKKQWHL 199
 R+ Y W LP

Sbjct: 183 RVCSYMFNWKLP 194

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3513> which encodes the amino acid sequence <SEQ ID 3514>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -5.15 Transmembrane 70 - 86 (65 - 88)
 INTEGRAL Likelihood = -4.09 Transmembrane 33 - 49 (29 - 49)
 INTEGRAL Likelihood = -2.13 Transmembrane 120 - 136 (119 - 137)
 INTEGRAL Likelihood = -0.43 Transmembrane 173 - 189 (172 - 189)

----- Final Results -----

-1266-

bacterial membrane --- Certainty=0.3060(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP:BAB05428 GB:AP001512 unknown conserved protein [Bacillus halodurans]
 Identities = 109/195 (55%), Positives = 137/195 (69%)

10 Query: 6 WEILNIIGTIAFALSGAIVAMEEEFDILGIFILGFVTAFGGGAIRNTLIGLPTEALWGQK 65
 W++LN+IGTIAFALSG IVAMEE+FD++G++ILGFVTAFGGGAIRN LIQ+P+ ALW Q
 Sbjct: 3 WDVNLVIGTIAFALSGVIVAMEEEDFDLMGVYILGFVTAFGGGAIRNLLIGVPVSALWEQG 62

15 Query: 66 PEFTCAFFAMVLIMLPKLMARGWVRAAVLTDAIGLAAFSVQCALHAVRLNQPLSAVIVT 125
 FT AF M + P L W++ +L DAIGLAAF++QGAL A ++ PLSAVIV
 Sbjct: 63 TLFTIAFIVMTIAFFLPNLWINHWLKFGLLFDAIGLAFAIQGALFATSMDHPLSAVIVA 122

20 Query: 126 AVLTGAGGGVVRDILAGRKPSVLRSEIYAGWSILAAIVLHFKLADSTIECYALVLLTTL 185
 A LTGAGGG+VRD+LA RKP VL EIY GW++LA + + I L++L+ L
 Sbjct: 123 AALTGAGGGIVRDM LARRKPLVLSKEIYIGWAMLAGAAIGLNIVSGPIGIGFLIILVVFL 182

Query: 186 RMIGNRKKWNLPKIK 200
 RM+ W LP K
 Sbjct: 183 RMLSVHYNWCLPHRK 197

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 133/200 (66%), Positives = 168/200 (83%)

30 Query: 3 MSIDIWNILSLVGTVAFASSGAIVAIEEEFDILGLFILGFVTAFGGGAIRNVLIGLPIET 62
 M+ID+W IL+++GT+AFA SGAIVA+EEEFDILG+FILGFVTAFGGGAIRN LIGLPIE
 Sbjct: 1 MTIDMWEILNIIGTIAFALSGAIVAMEEEFDILGIFILGFVTAFGGGAIRNTLIGLPIEA 60

35 Query: 63 LWSQGIAFYAAAAAILFIMIFPNLLSGKGRDAEVVSDAIGLAAFSVQCALYATQSHOPLS 122
 LW Q F A A++ IM+FP L++ A V++DAIGLAAFSVQCAL+A + +QPLS
 Sbjct: 61 LWGQKPEFTCAFFAMVLIMLPKLMARGWVRAAVLTDAIGLAAFSVQCALHAVRLNQPLS 120

40 Query: 123 AVIVA AVL TGAGGGIVRDVLAGRKPGVLRSEIYAGWSILVGIILYFKIAKTTTDDYLLVL 182
 AVIV AVL TGAGGG+VRD+LAGRKP VLRSEIYAGWSIL I+L+FK+A +T + Y LV+
 Sbjct: 121 AVIVTAVLTGAGGGVVRDILAGRKPSVLRSEIYAGWSILAAIVLHFKLADSTIECYALVV 180

Query: 183 VVTSRLMLGYKKQWHLPVVR 202
 ++T+LRM+G +K+W+LP ++
 Sbjct: 181 LLTTLRMIGNRKKWNLPKIK 200

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1133

A DNA sequence (GBSx1209) was identified in *S.galactiae* <SEQ ID 3515> which encodes the amino acid sequence <SEQ ID 3516>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2805(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9569> which encodes amino acid sequence <SEQ ID 9570> was also identified.

-1267-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB94816 GB:AJ245582 hypothetical protein [Streptococcus thermophilus]
Identities = 138/238 (57%), Positives = 184/238 (76%)

```

5   Query: 5   KKMIKLIAIDMDGTLNDEKKIPKENIQAIKEATQAGIKIVLCTGRPMGILPYFNELGL 64
      +   +KLIAIDMDGTLN +K+IPKENI+AI+EAT AGIKIVLCTGRP SGI+P+F +LGL
      Sbjct: 3   QNQVKLIAIDMDGTLNLSQKEIPKENIKAIQEATAAGIKIVLCTGRPRGIVPHFEKLG 62

10  Query: 65   TKEEYIIMNNGCSTYSTKDWQLIDSATLTHDELIFLEEVVKEFPNVCLTLTAENTFYAVG 124
      ++EE+IIMNNGCSTY TK+W L++S +L+ E+ L + ++FP V LT T E ++Y VG
      Sbjct: 63   SEEEFIIMNNGCSTYETKNWILLESESLRSEMEELLOQACEDFPGVALTFTGKESYYVVG 122

      Query: 125  EEVPEIVAYDADLVFTKAKSTSLDALRNQEEIVFQAMYMGLDADVTAFQEAEEALISKF 184
      EVPE+VAYDA VFT+AK+ SL+ + + +++FQAMYM + AFQ AV++ L +
15  Sbjct: 123  NEVPVLVAYDAGTVFTEAKARSLEEIFEEGQVIFQAMYMAESEPLDAFQNAVQDRLDQSY 182

      Query: 185  SGVRSQDYIYEIMPQGVTKARGLKSLIAKLGLDINQVMAIGDAPNDIELLDLVPSVA 242
      S VRSQ+YI+E+MPQG TKA GLK L KL ++ +Q+MA+GDA ND+E+L V SVA
20  Sbjct: 183  STVRSQEIYIFVMPQGATKASGLKHLAEKLDINRDQIMALGDAANDLEMLQFVGQSV 240

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3517> which encodes the amino acid sequence <SEQ ID 3518>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

25  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1468(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 152/270 (56%), Positives = 193/270 (71%)

35  Query: 6   KMIKLIADMDGTLNDEKKIPKENIQAIKEATQAGIKIVLCTGRPMGILPYFNELGLT 65
      +MI+LIAID+DGTLLN +K+IPKENI AI+EA Q+G+KIVLCTGRP SG PYF++LGLT
      Sbjct: 19  RMIQLIAIDLDGTLNLDKQIPKENITAIQEAAQSGLKIVLCTGRPQSGTRPYFDQLGLT 78

      Query: 66   KEEYIIMNNGCSTYSTKDWQLIDSATLTHDELIFLEEVVKEFPNVCLTLTAENTFYAVGE 125
      +EE++I+NNGCSTYS+ DWQL S L ++ LEE+ + FP++ LTLT EN + + E
40  Sbjct: 79   QEEFLIINNGCSTYSSPDWQLRHSKMLKVSDIELLELSQSFPDIYLTLEENDYLVLEE 138

      Query: 126  EVPEIVAYDADLVFTKAKSTSLDALRNQEEIVFQAMYMGLDADVTAFQEAEEALISKFS 185
      EVP++V D DLVFT K SL L + ++FQAMY+G A + AF+ AV L F
45  Sbjct: 139  EVPDLVQEDGDLVFTIIVKPVSLAELSDDTPRLIFQAMYLGEKAALDAFERAVRNQLSQSFH 198

      Query: 186  GVRSDQDYIYEIMPQGVTKARGLKSLIAKLGLDINQVMAIGDAPNDIELLDLVPSNVAMGN 245
      VRSQD I EI+PQGV+KA LK L+ LGL +QVMAIGDAPNDIE+L VAM N
50  Sbjct: 199  VVRSQDNILEILPQGVSKASALKELVEDLGLTADQVMAIGDAPNDIEMITYAGLGVAMEN 258

      Query: 246  ASDEIKSRCKYITVDNNKAGVAKAIYDYAL 275
      AS IK +T+ N+ AGVA+AI +AL
      Sbjct: 259  ASAAIKPLADKVTLTNDMAGVAQAIRQFAL 288

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1134

A DNA sequence (GBSx1210) was identified in *S.agalactiae* <SEQ ID 3519> which encodes the amino acid sequence <SEQ ID 3520>. Analysis of this protein sequence reveals the following:

Possible site: 18

-1268-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.43 Transmembrane 7 - 23 (7 - 23)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAA26954 GB:J04479 DNA polymerase I [Streptococcus pneumoniae]
 Identities = 655/879 (74%), Positives = 748/879 (84%), Gaps = 4/879 (0%)

Query: 3 NKNKLLLDIGSSVAFRAFFALYNQIDRFKNNSGLHTNAIYGFHMLNHLGRVQPSHILV 62
 +K KLLLDIGSSVAFRAFFALY Q+DRFKN +GLHTNAIYGF LML+H+L RV+PSHILV

15 Sbjct: 2 DKKKLLLDIGSSVAFRAFFALYQQLDRFKNAAGLHTNAIYGFQMLSHLLERVEPSHILV 61

Query: 63 AFDAGKTTFRTEMYADYKGGRAKTPDEFREQFPPIRQQLDVLGSIKHYELEHYEADDIIGT 122
 AFDAGKTTFRTEMYADYKGGRAKTPDEFREQFP+IR+ LD +GI+HYEL YEADDIIGT

20 Sbjct: 62 AFDAGKTTFRTEMYADYKGGRAKTPDEFREQFPFIRELLDHMGIRHYELAQYEADDIIGT 121

Query: 123 LAKQAEASNEHFDITVSGDKDLIQLTDTNIVVEISKKGVAEFEEFTPAYLMEKMGITPS 182
 L K AE + FDTIT+VSGDKDLIQLTD +TVVEISKKGVAEFE FTP YLME+MG+TP+

25 Sbjct: 122 LDKLAE--QDGFDTITVSGDKDLIQLTDEHTVVEISKKGVAEFEAFTPDYLMEMGLTPA 179

Query: 183 QFIDLKALMGDKSDNIPGVTKIGECTGLKLLSEYGSLEGIVENIEAMKQSKMKENLINDK 242
 QFIDLKALMGDKSDNIPGVTK+GEKTG+KLL E+GSLEGIVENI+ MK SKMKENLINDK

30 Sbjct: 180 QFIDLKALMGDKSDNIPGVTKVGEKTGIKLLLEHGSLEGIVENIDGMKTSKMKENLINDK 239

Query: 243 EQAFLSKTLATINIASPITIGLEDILYSGPQDIKALSQFYDEMDFKQFKAALGEETSQED 302
 EQAFLSKTLATI+ +PI IGLED++YSGP D++ L +FYDEM FKQ K AL ++

35 Sbjct: 240 EQAFLSKTLATIDTKAPIAIGLEDILVYSGP-DVENLGKFYDEMGMFKQLKQALNMSSADVA 298

Query: 303 FEVDFTVEVEQLKTEMFSDNDFYFEMLGDNVHVEDLIGIAWNSDTIYATSNVSLLEAL 362
 +DFT V+Q+ +M S+ ++FE+ G+NYH ++L+G AW D +YAT + LLQ+ +

40 Sbjct: 299 EGLDFTIVDQISQDMLSEESIFHFELFGENYHTDNLVGFPAWSCGDQLYATDKLELLQDPI 358

Query: 363 FKKALSKP-IKTYDFKRKSVLLNRFNIDLPAPFDTRLAKYLLSTTEDNLVSTIARLYTN 421
 FK L K ++ YDFK+ KVLL RF +DL PAPD RLAKYLLST EDN ++TIA LY

45 Sbjct: 359 FKDFLEKTSLRVYDFKKVKVLLQRFVLDLQAPAFDIRLAKYLLSTVEDNEIATIASLYGQ 418

Query: 422 LPLDITDAVYKKGAKRAIPEKTRFLEHLAKVKVLVDSEANIMQOLKANEQEELLFEME 481
 L D+ YGKG K+AIPE+ +FLEHLA K+ VLV++E ++++L N Q ELL++MEQ

50 Sbjct: 419 TYLVDDETFYGKGVKKAIPEREKFLLEHLACKLAVLVETEPILLEKLENGQLELLYDMEQ 478

Query: 482 PLANVLAKMEIRGIKVKKNTLNEMAIENQKVIETLTQEIYELAGQEFNINSPKQLGKLLF 541
 PLA VLAKMEI GI VKK TL EM EN+ VIE LTQEIYELAG+EFN+NSPKQLG LLF

55 Sbjct: 479 PLAFVLAKMEIAGIVVKKETLLEMQAENELVIEKLTQEIYELAGEEFNVNNSPKQLGVLLF 538

Query: 542 ETGLPLVEMTKKTKGYSTAVDVLERLAPISPLVTKILEYRQITKLQSTYIIGLQDYILE 601
 E LGLP+E TKKTKGYSTAVDVLERLAPI+P+V KIL+YRQI K+QSTY+IGLQD+IL

60 Sbjct: 539 EKLGLPLEYTKKTKGYSTAVDVLERLAPIPIVKKILDYRQIAKTQSTYVIGLQDWILA 598

Query: 602 DGKIHTRYVDLTQTGRLLSSDPNLQNIIPVRLEQGRLLIRKAFVPSEDNAVLLSSDYSQIE 661
 DGKIHTRYVDLTQTGRLLSS DPNLQNIIP RLEQGRLLIRKAFVP +++VLLSSDYSQIE

65 Sbjct: 599 DGKIHTRYVDLTQTGRLLSSVDPNLQNIIPARLEQGRLLIRKAFVPEWEDSVLLSSDYSQIE 658

Query: 662 LRVLAHISKDEHLIAAFKEGADIHTSTAMRVFGIEKPENVTNDRRNAKAVNFGIVYGIS 721
 LRVLAHISKDEHLI AF+EGADIHTSTAMRVFGIE+P+NVT NDRRNAKAVNFG+VYGIS

70 Sbjct: 659 LRVLAHISKDEHLIAKAFQEGADIHTSTAMRVFGIERPDNVTANDRRNAKAVNFGVYGIS 718

Query: 722 DFGLSHNLGIPRKLAKQYIDTYFERYPGIKNYMETVVREAKDKGYVETLFHRRRSLPDIN 781
 DFGLS+NLGI RK AK YIDTYFER+PGIKNYM+ VVREA+DKGYVETLF RRR LPDIN

75 Sbjct: 719 DFGLSNNLGISRKEAKAYIDTYFERFPGIKNYMDEVVREARDKGYVETLFRRRRBLPDIN 778

Query: 782 SRNFNIRQFAERTAINSPIQSSAADILKIAMINLDRVLDKGGYKSKMLLQVHDEIVLEVP 841
 SRNFNIR FAE TAINSPIQSSAADILKIAM LD+ L GGY++KMLLQVHDEIVLEVP

80 Sbjct: 779 SRNFNIRGFAERTAINSPIQSSAADILKIAMIQDLKALVAGGYQTKMLLQVHDEIVLEVP 838

-1269-

Query: 842 NEEIGAIRELVTKTMEASISLSVPLIADENAGETWYEAK 880
 E+ +++LV +TME AI LSVPLIADEN G TWYEAK
 Sbjct: 839 KSELVEMKKLVKQTMEEAIQLSVPLIADENEGATWYEAK 877

5

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3521> which encodes the amino acid sequence <SEQ ID 3522>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

10

INTEGRAL Likelihood = -0.43 Transmembrane 7 - 23 (7 - 23)

----- Final Results -----

bacterial membrane --- Certainty=0.1171(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 665/881 (75%), Positives = 761/881 (85%), Gaps = 2/881 (0%)

20

Query: 1 MTNKNKLLLLIDGSSVAFRAFFALYNQIDRFKNNSGLHTNAIYGFHMLNLHILGRVQPSHI 60
 M NKNKLLLLIDGSSVAFRAFFALYNQIDRFKN+SGLHTNAIYGFHML+H++ RVQP+H+

Sbjct: 1 MENKNKLLLLIDGSSVAFRAFFALYNQIDRFKNHSGHLHTNAIYGFHMLDHMMKRVQPIHV 60

25

Query: 61 LVAFDAGKTTFRTEMYADYKGGRAKTPDEFREQFPYIRQQLDVLGIKHYELEHYEADDII 120
 LVAFDAGKTTFRTEMYADYK GRAKTP+EFREQFPYIR+ L LGI +YELEHYEADDII

Sbjct: 61 LVAFDAGKTTFRTEMYADYKAGRAKTPDEFREQFPYIREMLTALGIAYYELEHYEADDII 120

30

Query: 121 GTLAKQAEASNEHFDITVVS GDKDLIQLTDNTVVEISKKGVAEFEEFTPAYLMEKMGIT 180
 GTL K AE + FD+T+VSGDKDLIQLTD NTVVEISKKGVAEFEEFTPAYLMEKMG+T

Sbjct: 121 GTLDRMAERTEVPFVDITVVS GDKDLIQLTDNTVVEISKKGVAEFEEFTPAYLMEKMGIT 180

35

Query: 181 PSQFIDLKALMGDKSDNIPGVTKIGEKTGLKLLSEYGSLEGIYENIEMKQSKMENLIN 240
 P+QFIDLKALMGDKSDNIPGVTKIGEKTGLKLL E+GSLEGIYE+I+ K SKMENLIN

Sbjct: 181 PNQFIDLKALMGDKSDNIPGVTKIGEKTGLKLLHEFGSLEGIYEHIDGFKTSKMENLIN 240

40

Query: 241 DKEQAFLSKTLATINIASPITIGLEDILYSGPDIKALSQFYDEMDFKQFKAALGEETSQ 300
 D++QAFLSKTLATIN ASPITIGL+DI+Y+GP D+ +LSQFYDEMDF Q K L + Q

Sbjct: 241 DRDQAFLSKTLATINTASPITIGLDDIVNGP-DVASLSQFYDEMDFVQLKKGLASQMPQ 299

45

Query: 360 EALFKKALS KPIKTYDFKRSKVLNRFNIDLPEPAFDTRLAKYLLSTTEDNLVSTIARLY 419
 FK+ KPI TYDFKRSKVLN+ I+L P++D RLA YLLST EDN +STIAR++

Sbjct: 360 TDSFKQVFQKPIATYDFKRSKVLNLSHLGIELVAPSYDARLANLYLLSTVEDNELSTIARIF 419

50

Query: 420 TNLPLDTDDAVYKGAKRAIPEKTRFLEHLAKKVKVLVDSEANIMQQLKANEQEELLFEM 479
 T++ L+ DD VYKGAKRA+P+K LEHLA+KVKVL+DS++ ++ +L A+EQ +L +

Sbjct: 420 TDISLEEDDTVYKGAKRAVDPKDVILLEHLARKVKVLDSKQMLDKLTAHEQLDLYQNI 479

55

Query: 480 EQPLANVLAKMEIRGIKVKNTLNEMAIENQKVIETLTQEIYELAGQEFNINSPKQLGKL 539
 E PLANVLAKMEI GIKV + TL +MA +N+ +IE LTQEIY++AQEFNINSPKQLG +

Sbjct: 480 ELPLANVLAKMEIEGIKVNRLQDMAEQNKVIEALTOEIYDMAGQEFNINSPKQLGSI 539

60

Query: 600 LEDGKIHTRYVQDLTQTGRLLSSDPNLQNIPIRLEQQRLIRKAFVPSEDNAVLLSSDYSQ 659
 L DGKIHTRYVQDLTQTGRLLSS DPNLQNIPIRLEQQRLIRKAF PS ++AVLLSSDYSQ

Sbjct: 600 LADGKIHTRYVQDLTQTGRLLSSVDPNLQNIPIRLEQQRLIRKAFVPSHEDAVLLSSDYSQ 659

65

Query: 660 IELRVLAHISKDEHLIAAFKEGADIHTSTAMRVFGIEKPENVTPNDRRNAKAVNFGIVYG 719
 IELRVLAHIS DEHLIAAF EGADIHTSTAMRVFGI++ +VT NDRRNAKAVNFGIVYG

-1270-

Sbjct: 660 IELRVLAHISGDEHLIAAFNEGADIHTSTAMRVFGIDRAADVTDANDRRNAKAVNFGIVYG 719

Query: 720 ISDFGLSHNLGIPRKLAKQYIDTYFERYPGIKNYMETVVREAKDKGYVETLPHRRRSLPD 779
ISDFGLS+NLGI RK AK YIDTYFERYPGIK YME VVREAKDKGYVETLF RRR LPD

5 Sbjct: 720 ISDFGLSNLGIITRKQAKSYIDTYFERYPGIKAYMENNVREAKDKGYVETLTKRRRELDP 779

Query: 780 INSRNFNIRQFAERTAINSPIQGSAADILKIAMINLDRVLDKGGYKSKMLLQVHDEIVLE 839
INSRNFN+R FAERTAINSPIQGSAADILKIAMINLD+ L GG+++KMLLQVHDEIVLE

10 Sbjct: 780 INSRNFNVRSAERTAINSPIQGSAADILKIAMINLDKALQAGGFRKMLLQVHDEIVLE 839

Query: 840 VPNEEIGAIRELVTKTMEASISLSVPLIADENAGETWYEAK 880
VPN+E+ AI++LV TME+A+ L+VPL DE+ G +WYEAK

Sbjct: 840 VPNDeltaIKLVKDTMEAAVDLAVPLCVDESTGHSWYEAK 880

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1135

A DNA sequence (GBSx1211) was identified in *S. agalactiae* <SEQ ID 3523> which encodes the amino acid sequence <SEQ ID 3524>. Analysis of this protein sequence reveals the following:

20 Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.1880(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9571> which encodes amino acid sequence <SEQ ID 9572> was also identified.

- 30 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05860 GB:AP001514 unknown conserved protein [Bacillus halodurans]
Identities = 72/134 (53%), Positives = 94/134 (69%), Gaps = 3/134 (2%)

35 Query: 17 NPSDFMLKKNYLTKAKTIAVVGSLDRQETAAYQVSKIMQEAGYQIIPVNPKNAGQKILGQM 76
NPSD +K L +AK IAVVGSL + +Y VS MQ AGY+IIPVNP ++LG+

Sbjct: 4 NPSDEKIKQILQEAKRIAVVGSLSGNPDRSYMVSAMQHAGYEIIPVNP--TVDEVLEGEK 61

Query: 77 TYASLKDVTEHIDIVNIFRRSEYLPDIAREFLEVADIFWAQLGLESQEAETILKQAGHK 136
SL+D+ +DIVN+FRRSE+LPD+ARE +E+ A +FWAQLGLE++EA L+Q G

40 Sbjct: 62 AVPSLQDIEGAVDIVNVFRSEHLDPVARETVEIGAPVFWAQLGLENKEAYDYLQQHGVT 121

Query: 137 QIVMNKCLKVECCQK 150
I MN+C+KVE K

45 Sbjct: 122 SI-MNRICKVEHAK 134

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3525> which encodes the amino acid sequence <SEQ ID 3526>. Analysis of this protein sequence reveals the following:

50 Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.0837(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 87/141 (61%), Positives = 114/141 (80%)

-1271-

Query: 11 M V Y H F Q N P S D F M L K N Y L T K A K T I A V V G L S D R Q E T A A Y Q V S K I M Q E A G Y Q I I P V N P K N A G Q 70
 ++Y FQNP S+ +L K Y L A K T I A V V G L S D R ++T A A Y V + K M Q Y + I I P V N P K A G Q
 Sbjct: 1 V I Y S F Q N P S E D V L K A Y L E S A K T I A V V G L S D R K D T A A Y G V A K F M Q A M D Y R I I P V N P K L A G Q 60

Query: 71 K I L G Q M I Y A S L K D V T E H I D I V N I F R R S E Y L P D I A R E F L E V D A D I F W A Q L G L E S Q E A E T I L 130
 I L G + Y A S + K + + D I V + + F R R S E + L P + + A R + F L A + F W A Q L G L E + Q E A + T I L
 Sbjct: 61 L I L G E K V Y A S I K A I P F E V D I V D V F R R S E F L P E V A R D F L A G Q A K V F W A Q L G L E N Q E A Q T I L 120

Query: 131 K Q A G H K Q I V M N K C L K V E C Q K L 151
 + A G + I V M N + C L K + + + L
 Sbjct: 121 R S A G K E A I V M N R C L K I D Y L Q L 141

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1136

A DNA sequence (GBSx1212) was identified in *S. agalactiae* <SEQ ID 3527> which encodes the amino acid sequence <SEQ ID 3528>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3367(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9573> which encodes amino acid sequence <SEQ ID 9574> was also identified.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3529> which encodes the amino acid sequence <SEQ ID 3530>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4960(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 113/151 (74%), Positives = 133/151 (87%), Gaps = 1/151 (0%)

Query: 7 M D S H S H G H R P L D A Y E N V L E H L R E K R I R I T E T R K A I I S Y M V N S R E H P S A E K I Y N D L L P E Y P 66
 M D H S H + L D A Y E N V L E H L R E K I R I T E T R K A I I S Y M + S E H P S A + K I Y D L P + P
 Sbjct: 1 M D I H S H - Q Q A L D A Y E N V L E H L R E K H I R I T E T R K A I I S Y M I Q S T E H P S A D K I Y R D L Q P N F P 59

Query: 67 N M S L A T V Y N N L K V L V D E G F V T E L K L C N Y S T T Y Y D F M G H Q H L N I A C E D C G K I V D F V D V D L L 126
 N M S L A T V Y N N L K V L V D E G F V + E L K + N T T Y Y D F M G H Q H + N + C E C G K I D F + D V D + +
 Sbjct: 60 N M S L A T V Y N N L K V L V D E G F V S E L K I S N D L T T Y Y D F M G H Q H V N V V C R I C G K I A D F M D V D V M 119

Query: 127 D I S R E A H Q Q T G F E V T R V Q L V A Y G I C P E C Q R K 157
 D I + + E A H + Q T G + + V T R + + + A Y G I C P + C Q K
 Sbjct: 120 D I A K E A H E Q T G Y K V T R I P V I A Y G I C P D C Q A K 150

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1272-

Example 1137

A DNA sequence (GBSx1213) was identified in *S.galactiae* <SEQ ID 3531> which encodes the amino acid sequence <SEQ ID 3532>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -2.13    Transmembrane    16 - 32 ( 14 - 32)
    INTEGRAL    Likelihood = -1.81    Transmembrane    496 - 512 ( 496 - 515)

----- Final Results -----
10      bacterial membrane --- Certainty=0.1850(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAA06650 GB:AJ005645 sdrc [Staphylococcus aureus]
    Identities = 41/146 (28%), Positives = 63/146 (43%), Gaps = 13/146 (8%)

Query: 4   SQYNKWSIRRLKVGAAASVMIASGSIVALGQSHIVSAD----EMSQPKTTITAPTANTSTN 59
      ++ NK+SIR+ VG AS+++ + I L      +A+    E++Q K   TAP+ N +T
20  Sbjct: 16 NRLNKFSIRKYSVGTASILVGTTLIFGLSGHEAKAAEHTNGELNQSKNETTAPSENKTT- 74

Query: 60  VESSTDKALSKVTTMETTSSEMPK--MQNMAKVEKTS SDKPMMVATSVRKMMATPTPVAMT- 116
      D   K T   +++ PK M + A V++TS      +      T T   T
25  Sbjct: 75 --KKVDSRQLKDNTQTATADQPKVMSDSATVKETSSNMQSPQNATANQSTTKTSNVTTN 132

Query: 117 ---KTTTSVDEVKKSTD TAFKQTV DVP 139
      TT +E KS T K   P
30  Sbjct: 133 DKSSTTYSNETDKSNLTQAKDVSTTP 158

```

30 No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8735> and protein <SEQ ID 8736> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 7
McG: Discrim Score:      -0.92
35  GvH: Signal Score (-7.5): -2.48
    Possible site: 39
    >>> Seems to have no N-terminal signal sequence
    ALOM program    count: 2 value: -2.13 threshold: 0.0
    INTEGRAL    Likelihood = -2.13    Transmembrane    16 - 32 ( 14 - 32)
40  INTEGRAL    Likelihood = -1.81    Transmembrane    496 - 512 ( 496 - 515)
    PERIPHERAL Likelihood = 7.96      402
    modified ALOM score: 0.93

*** Reasoning Step: 3
45 ----- Final Results -----
      bacterial membrane --- Certainty=0.1850(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 485-489

```

The protein has homology with the following sequences in the databases:

```

D|5981|5780 leukotoxin > Insert characterized
55  SP|P16462|HLYA_ACTAC LEUKOTOXIN. > Edit characterized
    GP|141834|gb|AA21922.1||M27399 leukotoxin (Lta) {Actinobacillus actinomycetemcomitans}
    Insert characterized

Query: 210 VSLNGNTTGKEGQALLDQI|AND---KHSYQATIRVYGAKDGKVDLKNMISPKMVTINIP 266
      ++ NG+   + G+A +D +K +   KHS + T ++   G +DL +   +T P
60

```

-1273-

Sbjct: 488 ITRNGDRI-QSGKAYVDYLKKGEELAKHSDKFTKQILDPIKGNIDLSGIKGSTTLTFLNP 546

Query: 267 HITDMEVKNSLKMFAFKEKV-DVPAKYVSAKAKG-PFLAGVNE--TIPYEAFGGDCMLT 322
+T E + + + E + ++ K + K KG P GV + + A D +

5 Sbjct: 547 LLTAGKEERKTRQSGKYEFITELKVKGRTDWKVKGVPNSNGVYDFSNIQHAVTRDNKVL 606

Query: 323 RLILKASEGAKWSDNGVDKNSPLL-----PLKDLTKGKYFYQVSLNGNTAGKKQALLD 376
L A+ GAK V S ++ + D +KG+ ++++G A K GQ ++

10 Sbjct: 607 EARLIANLGAKDDYVFGSGSTIVNAGDGYDVEDYSGKRTG-ALTIDGRNATKAGQYKVE 665

Query: 377 QIKANGSHTYQATITITYGTDGKV 400
+ +G+ Q T++ TK GKV

Sbjct: 666 R-DLSGTQVLQETVSKQETKRGKV 688

- 15 SEQ ID 3532 (GBS1) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 3; MW 78kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 3; MW 53kDa).

The His-fusion protein was purified as shown in Figure 189, lane 5.

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.
- 20

Example 1138

A DNA sequence (GBSx1214) was identified in *S.agalactiae* <SEQ ID 3533> which encodes the amino acid sequence <SEQ ID 3534>. This protein is predicted to be response regulator (regX3). Analysis of this protein sequence reveals the following:

- 25 Possible site: 32
>>> Seems to have no N-terminal signal sequence
- Final Results -----
- 30 bacterial cytoplasm --- Certainty=0.3585(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 35 >GP:CAB54578 GB:AJ006397 response regulator [Streptococcus pneumoniae]
Identities = 143/228 (62%), Positives = 183/228 (79%), Gaps = 1/228 (0%)
- Query: 1 MTQKLLLVDDDEFIIDINRRYLEQAGYEVSVAADGIEALKEVDENRFDLIISDIMPKMD 60
M + +LLVDD E I DI++RYL QAGY+V VA DG+EAL+ + DLII+D+MMP+MD
- 40 Sbjct: 1 MGKTILLVDDVEITDIHQRYLIQAGYQVLVAHDGLEALELFKKKPIDLIITDVMMPMD 60
- Query: 61 GYDFISEVLVREPNOFFLFITAKVSEPDKIYSLSMGADDFISKPPSPRELVLRVKNILRR 120
GYD ISEV P QPFLFITAK SE DKYI LS+GADDFI+KPPSPRELVLRV NILRR
- Sbjct: 61 GYDLISEVQYLSPEQPFLFITAKTSEQDKIYGLSLGADDFIAKPPSPRELVLRVHILRR 120
- 45 Query: 121 IYGNHQQSEVLTIGDLVIDQKQRLVMVDCNTISLTNKSFDLLWILANHLNRVFSKTELYE 180
++ ++E++++G+L ++ V + + LT KSF+LLWILA++ RVFSKT+LYE
- Sbjct: 121 LH-RGGETELISLGNLKMNHSSHEVQIGEMLDLTVKSFELLWILASNPERVFSKTDLYE 179
- Query: 181 RVWGEEFLDDTNTNLNVHIALRNDLAKFSTDNTPTIKTVWGLGYKLEE 228
++W E+++DDTNTNLNVHIALR +LAK+S+D TPTIKTVWGLGYK+E+
- 50 Sbjct: 180 KIWKEDYVDDTNTNLNVHIALRQELAKYSSDQPTIKTVWGLGYKIEK 227

There is also homology to SEQ ID 1182.

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.
- 55

-1274-

Example 1139

A DNA sequence (GBSx1215) was identified in *S.agalactiae* <SEQ ID 3535> which encodes the amino acid sequence <SEQ ID 3536>. This protein is predicted to be histidine kinase (resE). Analysis of this protein sequence reveals the following:

```

5   Possible site: 25
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -9.13    Transmembrane  42 - 58 ( 33 - 65)
      INTEGRAL    Likelihood = -7.54    Transmembrane   7 - 23 (  3 - 29)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.4652(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15  The protein has homology with the following sequences in the GENPEPT database.
   >GP:CAB54579 GB:AJ006397 histidine kinase [Streptococcus pneumoniae]
      Identities = 190/343 (55%), Positives = 249/343 (72%)

20  Query: 1   MKLKYYIVIGYLISMLITVAGVFFGLNHMLIETRGVYYILSVTIIACIVGGIVNLFLLSS 60
      Sbjct: 1   MKLK YI++GY+IS L+T+ VF+ + MLI +Y++L +TI+A +VG ++LFLL
      Sbjct: 1   MKLKSYILVGYIISTLLTILVVFVAVQKMLIARKEIYFLLGMTIVASLVCAGISLFLLLP 60

      Query: 61  VFTSLKKLKQKMKDISQRCFDTKAQICSPQEFKDLEAFNQMSSELESTFKSLNESEREK 120
      Sbjct: 61  VFTSL KLK+ K ++ + F + ++ P EF+ L FN+MS +L+ +F SL ESEREK

25  Sbjct: 61  VFTSLGKLKEHAKRVAADKFPNLEVQGPVEFQQLGQTFNEMSHDLQVSFDSLEESEREK 120

      Query: 121 TMMIAQLSHDIKTPITSIQSTVEGILDGIISEEEVNYLNTISRQTNRLNHLVEELSFT 180
      Sbjct: 121 +MIAQLSHDIKTPITSIQ+TVEGILDGII E E +YL TI RQT RLN LVEEL+F+T
      Sbjct: 121 GLMIAQLSHDIKTPITSIQATVEGILDGIIKESEQAHYLATIGRQTERLNKLVEELNFLT 180

30  Query: 181 LETMSDTAEPHKETIYLDKLLIDILSEFQLVFEKENRQVMIDVAPDVSKLSSQYDKLSR 240
      Sbjct: 181 L T + E +++I+LDKLLI+ +SEFQ + E+E R V + V P+ +++ Y KLSR
      Sbjct: 181 LNTARNQVETTSKDSIFLDKLLIECMSRFQFLIEQERRDVHLQVIPESARIEGDYAKLSR 240

35  Query: 241 ILLNLISNAVKYSDPGSPLTIKAYSNRQDIVIDIIDQGYGIKDEDLASIFNRLYRVESSR 300
      Sbjct: 241 IL+NL+ NA KYS PG+ L + A + + I + D+G GI EDL +IF RLYRVE+SR
      Sbjct: 241 ILVNLVDNAFKYSAPGKLEVVAKLEKDQLSISVTDEGQGIAPEDLENIFKRLYRVETSR 300

      Query: 301 NMKTGGHGLGLYIARQLAHQLNGDILVESQYQKGSKFSLVLKL 343
      Sbjct: 301 NMKTGGHGLGL IAR+LAHQL G+I V SQY GS F+LVL L
      Sbjct: 301 NMKTGGHGLGLAIARELAHQLGGEITVSSQYGLGSTFTLVNL 343

```

There is also homology to SEQ ID 1178.

A related GBS gene <SEQ ID 8737> and protein <SEQ ID 8738> were also identified. Analysis of this protein sequence reveals the following:

```

45  Lipop: Possible site: -1   Crend: 3
      McG: Discrim Score:      8.67
      GvH: Signal Score (-7.5): -5.75
      Possible site: 25
50  >>> Seems to have an uncleavable N-term signal seq
      ALOM program  count: 2 value: -9.13 threshold: 0.0
      INTEGRAL    Likelihood = -9.13    Transmembrane  42 - 58 ( 33 - 65)
      INTEGRAL    Likelihood = -7.54    Transmembrane   7 - 23 (  3 - 29)
      PERIPHERAL  Likelihood =  3.92      196
55  modified ALOM score:  2.33

      *** Reasoning Step: 3

      ----- Final Results -----
60  bacterial membrane --- Certainty=0.4652(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```


-1276-

Possible site: 19

>>> Seems to have no N-terminal signal sequence

5	INTEGRAL	Likelihood = -7.70	Transmembrane	125 - 141 (110 - 155)
	INTEGRAL	Likelihood = -7.59	Transmembrane	38 - 54 (36 - 56)
	INTEGRAL	Likelihood = -6.48	Transmembrane	146 - 162 (143 - 174)
	INTEGRAL	Likelihood = -5.57	Transmembrane	72 - 88 (63 - 93)
	INTEGRAL	Likelihood = -1.33	Transmembrane	229 - 245 (227 - 245)

-1277-

An alignment of the GAS and GBS proteins is shown below.

Identities = 87/254 (34%), Positives = 137/254 (53%), Gaps = 10/254 (3%)

```

5  Query: 16 MTNSEIKNEAKTILSNLQGNQLFLLPILLSIITLYISFYQYN-----NMTLLDFFVPL 70
      M+  IK +A+ L NL GK LFL+P LL + I + Y ++L + PL
      Sbjct: 1 MSIKAIKGQARDTLKNLSGKYLFLIPTLLFMFHFGEIHQGYVLSSGIEVSLAASYFPL 60

10 Query: 71 PVYFFYTLPFIISVSFVMLDVVKQKLNVRFSNTYVFSSHIFWKLLSVLVKGLILSFFY 130
      + +LFI+S SF M+DVV++ + V F+++T FS F LL + + K L +
      Sbjct: 61 LLGLILSLFILSASFTMIDVVRHFRQKVSFAESTTAFSKEFFGNLLVLAITKWLFLLIWS 120

15 Query: 131 LLSTFGLLIISFRLL-----SGNLILAPVLIVVSSLITTKAVIKLVQQYYSYSISTL 185
      L+ FGL I +S L + +++ + ++ ++++ + +YY+YS+S
      Sbjct: 121 LIWFFGLFIFLSGLSAPLVNAKSGSSTVISLIFLLFGAVLSLIGFGIYINRYAYSLSSEY 180

20 Query: 186 VFYTOLESGNYEGPSKVLVASRELMNGNKLRLFLDLFSFIGWQFLTIFSFGLVYIYLLPY 245
      + Y +++ G Y G V+ S +M G K +LF L LSF GW L I +FGL+ IYLLPY
      Sbjct: 181 LLYDEVKEGTYLGAIAVIETSVAMMKGYKWLFFLQLSFTGWFLNIVTFGLLNIYLLPY 240

25 Query: 246 QTTARLIFYRNITK 259
      TTA +IFY + K
      Sbjct: 241 FTTANVIFYDQLKK 254

```

A related GBS gene <SEQ ID 8739> and protein <SEQ ID 8740> were also identified. Analysis of this protein sequence reveals the following:

```

30 Lipop: Possible site: -1  Crend: 4
    McG: Discrim Score: -11.32
    GvH: Signal Score (-7.5): -5.39
        Possible site: 19
    >>> Seems to have no N-terminal signal sequence
    ALOM program count: 5 value: -7.70 threshold: 0.0
        INTEGRAL Likelihood = -7.70 Transmembrane 125 - 141 ( 110 - 155)
        INTEGRAL Likelihood = -7.59 Transmembrane 38 - 54 ( 34 - 56)
        INTEGRAL Likelihood = -6.48 Transmembrane 146 - 162 ( 143 - 174)
35  INTEGRAL Likelihood = -5.57 Transmembrane 72 - 88 ( 63 - 93)
        INTEGRAL Likelihood = -1.33 Transmembrane 229 - 245 ( 227 - 245)
        PERIPHERAL Likelihood = 0.37 105
    modified ALOM score: 2.04

40 *** Reasoning Step: 3

----- Final Results -----
        bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

50 ORF00498(901 - 1071 of 1383)
    EGAD|19922|20421(155 - 211 of 226) hypothetical protein {Bacillus megaterium}
    GP|288299|emb|CAA79984.1||Z21972 ORF1 {Bacillus megaterium} PIR|S32215|S32215 hypothetical
    protein 1 - Bacillus megaterium
    %Match = 4.8
    %Identity = 36.8 %Similarity = 61.4
    Matches = 21 Mismatches = 22 Conservative Sub.s = 14

55 741 771 801 831 861 891 921 951
    LIIISSFRLLSGNLILAPVLIVVSSLITTKAVIKLVQQYYSYSISTLVFYTOLESGNYEGPSKVLVASRELMNGNKLRL
    : : | : | :
    GIPEVFAYIKDGKTSFKLIGASILQAFIFLWSLLLIIPGIIKAIAYSQQFFLLKDHPEYTVLEAITESKKRMKGLKWKY
60 110 120 130 140 150 160 170

    981 1011 1041 1071 1101 1131 1161 1191
    FLDDLFSFIGWQFLTIFSFGLVYIYLLPYQTTARLIFYRNITKNS*E*FLAIFVI*VLKRTYCLFDTDFRPKYPHSVDVQV
    ||: ||||| | :||:| :||:| | :

```

-1278-

FLMHLSPFIGWILCMFTLIGIGLLWLIPYAGTTTAAFYEEELIVPQEDIDDDQQIEG
 190 200 210 220

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1141

A DNA sequence (GBSx1217) was identified in *S.agalactiae* <SEQ ID 3541> which encodes the amino acid sequence <SEQ ID 3542>. This protein is predicted to be tRNA-guanine transglycosylase (tgt). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3706(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9577> which encodes amino acid sequence <SEQ ID 9578> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14731 GB:Z99118 tRNA-guanine transglycosylase [Bacillus subtilis]
 Identities = 269/377 (71%), Positives = 320/377 (84%)

Query: 12 MTDHPKIKYRLIKQEKHTGARLGEEIITPHGTFPTPMFMPVGTQATVKTQSPEELKEMGSGI 71
 M + PI+Y IK+ K TGARLG++ TPHG+F TP+FMPVGT ATVKT SPEELK M +GI
 Sbjct: 1 MAEQPIRYEFIKECKQTGARLGKVVHTPHGSFETPVFMPVGTLATVKTMSPEELKAMDAGI 60

Query: 72 ILSNTYHLWLRPGDELIKAGGLHKFMNWDQAILTDSGGFQVYSLADSRNITEEGVTFKN 131
 ILSNTYHLWLRPG +++ +AGGLHKFMNWD+AILTDSGGFQV+SL+ RNI EEGV F+N
 Sbjct: 61 ILSNTYHLWLRPGQDIVKEAGGLHKFMNWDRAILTDSGGFQVFSLSKFRNIEEGVHFRN 120

Query: 132 HLNGAKMFLSPEKAISIQNGLSDIMMSFDECPQFYQPYDYVKKSIERTSRWAERGLNAH 191
 HLNG K+FLSPEKA+ IQN LGSDIMM+FDECP + YDY+K+S+ERTSRWAER LNAH
 Sbjct: 121 HLNGDKLFLSPEKAMEIQNALGSDIMMAFDECPYPAYDYMKRSVERTSRWAERCLNAH 180

Query: 192 RRPDQGLFGIVQGAGFEDLRQASARDLVSMDFPGYSIGGLAVGETHDEMNAVLDFTVPM 251
 R +QGLFGIVQG +EDLR QSA+DL+S+DFPGY+IGGL+VGE D MN VL+FT P+
 Sbjct: 181 NRQDEQGLFGIVQGGYEEDLRTQSAKDLISLDFPGYAIGGLSVGEPKDVMMNRVLEFTTPL 240

Query: 252 LPNDKPRYLMGVGAPDSLIDAVIRGVDMFDCVLPTRIARNGTCMTSQGRLVVKNAKFAED 311
 LP DKPRYLMGVG+PD+LID IRGVDMFDCVLPTRIARNGT T++GRL +KNAKF D
 Sbjct: 241 LPKDKPRYLMGVGSPDALIDGAIRGVDMFDCVLPTRIARNGTVFTAEGRLNMKNAKFERD 300

Query: 312 FTPLDPNCDCYTCKNYTRAYIRHLKADETFGIRLTSYHNLYFLVNLMKDVRQAIMDDNL 371
 F P+D CDCYTCKNYTRAYIRHL++ +ETFG+RLT+YHNL+FL++LM+ VRQAI +D L
 Sbjct: 301 FRPIDEECDYTCKNYTRAYIRHLIRCNFTFGLRLTTYHNLHFLHLHLMQVQAIREDRL 360

Query: 372 LEFRQDFMERYGYGMNN 388
 +FR++F ERYGY N
 Sbjct: 361 GDFREEFFERYGYNKP 377

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3543> which encodes the amino acid sequence <SEQ ID 3544>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>

-1279-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 351/380 (92%), Positives = 368/380 (96%)

Query: 12 MTDHPKIYRLIKQEKTGARLGEIITPHGTFPTPMFMPVGTQATVKTQSPEELKEMGSGI 71
 MTD+PIKYRLIK EKHTGARLGEIITPHGTFPTPMFMPVGTQATVKTQSPEELK +GSGI
 10 Sbjct: 1 MTDYPIKYRLIKA EKHTGARLGEIITPHGTFPTPMFMPVGTQATVKTQSPEELKAIGSGI 60

Query: 72 ILSNTRYHLWLRPGDELIAGGLHKFMNWDQAILTDSGGFQVYSLADSRNITEEGVTFKN 131
 ILSNTRYHLWLRPGDELIA++GGLHKFMNWDQ ILTDSGGFQVYSLADSRNITEEGVTFKN
 15 Sbjct: 61 ILSNTRYHLWLRPGDELIARSGLHKFMNWDQPILTDSGGFQVYSLADSRNITEEGVTFKN 120

Query: 132 HLNGAKMFLSPEKAISTQNNLGS DIMMSFDECPQFYQPYDVVKKSIERTSRWAERGLNAH 191
 HLNG+KMFLSPEKAISTQNNLGS DIMMSFDECPQFYQPYDVVKKSIERTSRWAERGL AH
 20 Sbjct: 121 HLNGSKMFLSPEKAISTQNNLGS DIMMSFDECPQFYQPYDVVKKSIERTSRWAERGLKAH 180

Query: 192 RRPHDQGLFGIVQGAGFEDLRRQSA DLV+MDFFPGYSIGGLAVGETHDEMNAVLDFTVPM 251
 RRPHDQGLFGIVQGAGFEDLRRQSA DLV+MDFFPGYSIGGLAVGE+H+EMNAVLDFTV P+
 25 Sbjct: 181 RRPHDQGLFGIVQGAGFEDLRRQSAADLVAMDFPGYSIGGLAVGESHEEMNAVLDFTTPL 240

Query: 252 LPNDKPRYLMGVGAPDSLIDAVIRGVMFDCVLPTRIARNGTCMTSQGRLVVKNAKFAED 311
 LP +KPRYLMGVGAPDSLID VIRGVMFDCVLPTRIARNGTCMTS+GRLV+KNAKFAED
 30 Sbjct: 241 LPENKPRYLMGVGAPDSLIDGAVIRGVMFDCVLPTRIARNGTCMTSEGRVVIKNAKFAED 300

Query: 312 FTPLDPNCDCYTCKNYTRAYIRHLLKADETFGIRLTSYHNLYFLVNLMKDVRQAIMDDNL 371
 FTPLD +CDCYTC+NY+RAYIRHLLKADETFGIRLTSYHNLYFLVNLMK VRQAIMDDNL
 35 Sbjct: 301 FTPLDHDCCDCYTQNYSRAYIRHLLKADETFGIRLTSYHNLYFLVNLMKKVRQAIMDDNL 360

Query: 372 LEFRQDFMERYGYGMNRRNF 391
 LEFRQDF+ERYGY +NRNF
 40 Sbjct: 361 LEFRQDFLERYGYNKSRRNF 380

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1142

A DNA sequence (GBSx1218) was identified in *Sagalactiae* <SEQ ID 3545> which encodes the amino acid sequence <SEQ ID 3546>. Analysis of this protein sequence reveals the following:

40 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2479(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9303> which encodes amino acid sequence <SEQ ID 9304> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10795> which encodes amino acid sequence <SEQ ID 10796> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:CAB16256 GB:Z99164 hypothetical protein [Schizosaccharomyces
 pombel]
 Identities = 42/91 (46%), Positives = 62/91 (67%), Gaps = 3/91 (3%)

Query: 6 FGIGLDSSSR CYHYHTKLDIVALKCAVCQKYACYKCHDALEEHCF AA-TKSDETFP-VL 63
 +G +D+ +RC+HYH+K D+VAL+C C+K+YAC++CHD L H F K+ P V+
 Sbjct: 13 YGKLV DNETRCFHYHSKADVVALRCGQCEK FYACFQCHDELNTHPFLPWRKAKFHIPC VI 72

-1280-

Query: 64 CGSCRQMLTLKEYK-TGFCPYCRMLFNPNCQ 93
 CG+C+ LT++EY+ T C YC FNP C+
 Sbjct: 73 CGACKNSLTVEEYRSTVHCKYCNHFPNPKCK 103

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3547> which encodes the amino acid sequence <SEQ ID 3548>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2769(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 55/93 (59%), Positives = 62/93 (66%)

Query: 2 MQEYFGIGLDSSSRCYHYHTKLDIVALKCAVCQKYACIKCHDALEEHCFATKSDETFF 61
 M + FGI LD RC HYHT LDIV LKCA CQ YYACY CHD L +H F T ET P
 Sbjct: 1 MTDCFGIDLDQEYRCLHYHTPLDIVGLKCAQCQYYACYHCHDQLTDHAFVPTGHQETSP 60

Query: 62 VLCGSCRQMLTLKEYKTGFCPYCRMLFNPNCQR 94
 V+CG CR++L+ EY G CPYC+ FNP C R
 Sbjct: 61 VICGHCRKLLSRAEYGCGCCPYCQSPFPNACHR 93

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1143

A DNA sequence (GBSx1219) was identified in *S.agalactiae* <SEQ ID 3549> which encodes the amino acid sequence <SEQ ID 3550>. This protein is predicted to be transport protein. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.45 Transmembrane 300 - 316 (292 - 321)
 INTEGRAL Likelihood = -1.17 Transmembrane 265 - 281 (265 - 281)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10113> which encodes amino acid sequence <SEQ ID 10114> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF12002 GB:AE002075 transport protein, putative [Deinococcus radiodurans]
 Identities = 108/295 (36%), Positives = 174/295 (58%), Gaps = 4/295 (1%)

Query: 31 GAWINLVNPSQEESEQVADQFGIDIDDLRAPLDVEETSRISEDDYTLVIVDVPTYEERN 90
 G WI+ P+ EE +V+ + G+++D L+ PLD +E SR ED L+I+ +
 Sbjct: 21 GCWIDAAAPTTEELARVSRETGLELDYLYKPLDPDERSRFREREDGQLLIIMQTSYRLAED 80

Query: 91 NKSYMTIPMGIIVTDNAVITTC-LEHLTLFDHFYRRRVKNFYTFMKTRFVFQLLYRNAE 149
 + Y T+P+GI+ TD+ ++T C LE + V+ T K R QL RNA+
 Sbjct: 81 SDIPYDTPVLGILHTDHCLVTVCSEENPVVKDVVSGLVRRVSTVKKNRLTLQLFLRNAQ 140

-1281-

Query: 150 LYLQALRTIDRQSDKIEAQLESATRNEQLIDMMELEKSIVYLKASLKFNERTIVKKLTSST 209
 +L +R I+++ D IE ++E+ATRN +L+D+++LEKS+VY LK NE +++++
 Sbjct: 141 RFLIDVRQINKRVDAIEDKMNATRNRELLDLLKLEKSLVYFITGLKANEAMMERVKRDR 200

5 Query: 210 SSLKKYIEDEDLLEDTLIETQQAIEMANIYENVLNAMTETTASIIGNNQNTIMKTLALVT 269
 + Y ED +LL+D LIE QAIEMA+I N+L +M AS+I NN N ++K L + T
 Sbjct: 201 I-FEMYEDSELDDVLIENLQAIEMASIASNILTSMAGAFASVINNNVNQVVKVLTVT 259

10 Query: 270 MTLDIPTVIFSAYGMNFQNNWMLNGLAHGFIYVLLAFLMSSFVVFYFIRKKWF 324
 + + IPT++ +GMN + +P + +GF V+ +A ++S + F F R K F
 Sbjct: 260 ILVAIPTLVSGFFGMNVEG--LPFSDSPYGFVLVMTVAMGIASLLAFLFYRWKVF 312

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 715> which encodes the amino acid sequence <SEQ ID 716>. Analysis of this protein sequence reveals the following:

15 Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.81 Transmembrane 293 - 309 (288 - 311)
 INTEGRAL Likelihood = -1.28 Transmembrane 255 - 271 (255 - 271)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.4524(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 272/314 (86%), Positives = 296/314 (93%)

Query: 11 MKQMFLSTAIEFKEIETFEFGAWINLVNPSQEESQVADQFGIDIDDLRAPLDVEETSRI 70
 MKQMFLS+AIEFKEIETFEFGAWI LVNPSQEEs ++ADQF IDI DLRAPLDVEETSRI
 30 Sbjct: 1 MKQMFLSSAIEFKEIETFEFGAWIKLVNPSQEESMKIADQFNIDISDLRAPLDVEETSRI 60

Query: 71 SVEDDYTLVIVDVPTIYERNNKSYYMTIPMGIIIVTDNAVITTCLEHLTLFDHFYRRRVKN 130
 +VEDDYTL+IVDVP YEERNNKSYY+T+P+GIIVT+NAVITTCLEHLTLFDHF+ RRVKN
 Sbjct: 61 AVEDDYTLIIVDVPIYERNNKSYYITMPLGIIIVTENAVITTCLEHDMTLFDHFHNRVKN 120

35 Query: 131 FYTFMKTRFVFQQLYRNAELYLQALRTIDRQSDKIEAQLESATRNEQLIDMMELEKSIVY 190
 FYTFMKTRFVFQ+LYRNAEL+L ALRTIDRQS+++EAQLE+ATRNE+LIDMMELEKSIVY
 Sbjct: 121 FYTFMKTRFVFQQLYRNAELFLTAIRTIDRQSERLEAQLEATRNEELIDMMELEKSIVY 180

40 Query: 191 LKASLKFNERTIVKKLTSSTSSLKKYIEDEDLLEDTLIETQQAIEMANIYENVLNAMTETT 250
 LKASLKFNERTIVKKL+SSTSSLKKYIEDEDLLEDTLIETQQAIEMA IYENVLNAMTETT
 Sbjct: 181 LKASLKFNERTIVKKLSSSTSSLKKYIEDEDLLEDTLIETQQAIEMAGIYENVLNAMTETT 240

45 Query: 251 ASIIGNNQNTIMKTLALVTMTLDIPTVIFSAYGMNFQNNWMLNGLAHGFIYVLLAFLM 310
 ASII NNQNTIMKTLAL+TM LDIPTVIFSAYGMNFQNNW+PLNGL H F Y+ L+A L+
 Sbjct: 241 ASIINNQNNTIMKTLALMTMALDIPTVIFSAYGMNFQNNWLPNGLEHAFWYITLIAMLL 300

Query: 311 SSFVVFYFIRKKWF 324
 SSFVV YFIRKKWF
 50 Sbjct: 301 SSFVVIYFIRKKWF 314

SEQ ID 3550 (GBS257) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 3; MW 35kDa), in Figure 169 (lane 9 & 10; MW 50kDa) and in Figure 239 (lane 2; MW 50kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 6; MW 60kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1282-

Example 1144

A DNA sequence (GBSx1220) was identified in *S.galactiae* <SEQ ID 3551> which encodes the amino acid sequence <SEQ ID 3552>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 29
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -12.26    Transmembrane  158 - 174 ( 151 - 182)
      INTEGRAL    Likelihood = -6.37     Transmembrane   93 - 109 ( 91 - 111)
      INTEGRAL    Likelihood = -5.68     Transmembrane  188 - 204 ( 184 - 205)
10   INTEGRAL    Likelihood = -0.85     Transmembrane  118 - 134 ( 118 - 134)

      ----- Final Results -----
                bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
                bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
15                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3553> which encodes the amino acid sequence <SEQ ID 3554>. Analysis of this protein sequence reveals the following:

```

20   Possible site: 52
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -6.95     Transmembrane   92 - 108 ( 88 - 110)
      INTEGRAL    Likelihood = -6.69     Transmembrane  153 - 169 ( 151 - 177)
      INTEGRAL    Likelihood = -2.34     Transmembrane  183 - 199 ( 183 - 200)

25   ----- Final Results -----
                bacterial membrane --- Certainty=0.3781(Affirmative) < succ>
                bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

30 The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 135/217 (62%), Positives = 167/217 (76%), Gaps = 1/217 (0%)

35   Query: 1   MTLQDLTKKNQEFVHIATNQLLADGKSDAEIKAILEEHLPEIIDNQKKGITARSLLGAPT 60
      M LQ+LTTKKNQEF+H ATN+L+ DGKSD +IK ILEE +P I++NOKKG+TAR+LLG PT
      Sbjct: 1   MELQELTKKNQEFIHATNKLQDGKSDEDIKLILEEAI PAILENOKKGV TARNLLGTPT 60

      Query: 61  TWAASFTEPDKARVSVQKNTNPWLMWLDTSLLFLGLVTALNGLMLLFGQSNVNTGLIS 120
      WAASF++ P KA KNTNPWLMWLDTSLLF+G+V LNG+M F + TGLIS
40   Sbjct: 61  AWAASFSDPSQKA-AETDKNTNPWLMWLDTSLLFIGIVALNGINTFFNTNATVTGLIS 119

      Query: 121 ILTLGFGGGAAMYVTTYIYRHMKGPKSERPGWLKSFVALLVLMVWFALFAVPLL PAT 180
      +L LGFGGGA+MY TTY+IYRH+GK KS RP W K A L+L ML+W AL++ LP +
45   Sbjct: 120 LLALGFGGASMYATYYFIYRHLGKDKSLRPSWFKIIAALSLAMLIWIALYSATAFLPTS 179

      Query: 181 INPKLPEVVLFIIALASEFGLRFYLQKRYNIQSSMAPV 217
      +NP+LP + L II S LR+YLQKRYNIQ++M+PV
      Sbjct: 180 LNPQLPLALLIIGGVSLALRYYLQKRYNIQNTMSPV 216

```

50 A related GBS gene <SEQ ID 10787> and protein <SEQ ID 10788> were also identified. Analysis of this protein sequence reveals the following:

```

      Lipop: Possible site: -1    Crend: 8
      McG: Discrim Score:      -9.94
      GvH: Signal Score (-7.5): -3.66
55   Possible site: 29
   >>> Seems to have no N-terminal signal sequence
      ALOM program count: 4 value: -12.26 threshold: 0.0
      INTEGRAL    Likelihood = -12.26    Transmembrane  158 - 174 ( 151 - 182)

```


-1283-

INTEGRAL Likelihood = -6.37 Transmembrane 93 - 109 (91 - 111)
 INTEGRAL Likelihood = -5.68 Transmembrane 188 - 204 (184 - 205)
 INTEGRAL Likelihood = -0.85 Transmembrane 118 - 134 (118 - 134)
 PERIPHERAL Likelihood = 8.43 50
 modified ALOM score: 2.95
 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1145

A DNA sequence (GBSx1221) was identified in *Sagalactiae* <SEQ ID 3555> which encodes the amino acid sequence <SEQ ID 3556>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1348(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1146

A DNA sequence (GBSx1222) was identified in *Sagalactiae* <SEQ ID 3557> which encodes the amino acid sequence <SEQ ID 3558>. This protein is predicted to be excinuclease ABC (uvrA). Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10111> which encodes amino acid sequence <SEQ ID 10112> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC67271 GB:AF017113 excinuclease ABC subunit A [Bacillus subtilis]
 Identities = 642/940 (68%), Positives = 785/940 (83%), Gaps = 3/940 (0%)

Query: 9 DKLMIKGARAHNLKNISVDIPRDKLVVVVTGLSGSGKSSLAFTTIYAEGQRRYVESLSAYA 68
 D++ ++GARAHNLKNI V IPRD+LVVVVTGLSGSGKSSLAFTTIYAEGQRRYVESLSAYA
 Sbjct: 4 DRIEVKGARAHNLKNIDVTIPRDKLVVVVTGLSGSGKSSLAFTTIYAEGQRRYVESLSAYA 63

-1284-

Query: 69 RQFLGNMEKPDVDSIDGLSPAISIDQKTTSKNPRSTVGTVTETINDYLRLLYARVGTPPYCI 128
 RQFLG M+KPDVD+I+GLSPAISIDQKTT+NRSTVGTVTETI DYLRLLYARVG P+C
 Sbjct: 64 RQFLGQMDKPDVDAIEGLSPAISIDQKTTSRNPRSTVGTVTETIYDYLRLLYARVGKPHCP 123

5 Query: 129 NGHGAITASSVEQIVDKVLALPERTKMQLAPIIRRKKGQHKSTFEKIQKDGYYVRVRIDG 188
 IT+ ++EQ+VD++L PERTK+Q+LAPI+ +KG H E+I+K GYVRVRIDG
 Sbjct: 124 EHGIEITSQTIEQMVDRILEYPERTKLQVLAPIVSGRKGAVKVLQEQIRKQGYVRVRIDG 183

10 Query: 189 DIHDVTEVPPELSKSKMHNIDIVVDRLINKEGIRSRFLDSVEAALRLSDGYVVIDTMDGNE 248
 ++ ++++ EL K+K H+I++V+DR++ KEG+ +RL DS+E ALRL +G V+ID + E
 Sbjct: 184 EMAELSDDIELEKNKKHSIEVVDRIVVKEGVAARLSDSLETALRLGEGRVMIDVIGEEE 243

15 Query: 249 LLFSEHYSCPECGFTVPELEPRLFSFNAPFGSCPTCDGLGIKLEVDIDLVIPDRSKTLRE 308
 L+PSEH++CP CGF++ ELEPRLFSFN+PFG+CPTCDGLG+KLEVD DLVIP++ +L+E
 Sbjct: 244 LMFSEHHACPHCGFSIGELEPRLFSFNSPFGACPTCDGLGMKLEVDADLVIPNQLSLKE 303

20 Query: 309 GALVPWNPISSNYPTMLEQAMTQFGVMDTPEFKLSKAEQDLALYSGGEREFHFHYIND 368
 A+ PW PISS YYP +LE T +G+DMD P + L K + D LYSGG+ +F Y ND
 Sbjct: 304 NAVAPWTPPISSQYYPQLLEAVCTHYGIDMDVPVKDLPHQLDKVLYSGGDDLIFRYEND 363

25 Query: 369 FGGERNIDLFFEGVNNINRRYHETNSDYTRNVMREYMNELKCNCHGYRLNDQALCVRV 428
 FG R ++ FEGV+ NI RRY ET SD+ R M +YM++ C TC GYRL +AL V +
 Sbjct: 364 FGQIREGEIQFEGVLNRIERRYKETGSDFIREQMEQYMSQKSCPTCKGYRLKKEALAVLI 423

30 Query: 429 GGEGLNIGQVSDLSIADHLELLETLRLSSNEQLIARPIIKEIHDRLSFLNNVGLNYLNL 488
 +G +IG++++LS+AD L + L LS + IA I++EI +RLSFL+ VGL+YL L
 Sbjct: 424 ---DGRHIGKITELSVADALAFFKDLTLSEKDMQIANLILREIVERLSFLDKVGLDYLT 480

35 Query: 489 SRSAGTSLGGESQIRLATQIGSNLSGVLYVLDEPSIGLHQDNDRLIDSLKMRDLGNT 548
 SR+AGTSLSGGE+QRIRLATQIGS LSGVLY+LDEPSIGLHQDNDRLI +LK MRDLGNT
 Sbjct: 481 SRAAGTSLSGGEAQRIRLATQIGSRLSGVLYILDEPSIGLHQDNDRLISALKNMRDLGNT 540

40 Query: 549 LIVVEHDEDTMAADWLIDVGFGAGAFGGEIVASGTPKQVAKNTKSITGQYLSGKKVIPV 608
 LIVVEHDEDTMAAD+LID+GFGAG GG+++++GTP++V ++ S+TG YLSGKK IP+
 Sbjct: 541 LIVVEHDEDTMAADYLIDIGFGAGIHGGQVISAGTPBEVMEDPNSLTGSYLSGKKFIPL 600

45 Query: 609 PSERRVGNRFLLEIKGAAENNLQNLVDVKFPLGKFIATVGVSGSGKSTLINSILKKAQAQK 668
 P ERR +GR++EIKGA+ENNL+ ++ KPFLG F AVTGVSGSGKSTL+N IL KA+AQK
 Sbjct: 601 PPERRKPDGRYIEIKGASENNLKKVNAKFPPLGTFATVGVSGSGKSTLVNEILHKALAQK 660

50 Query: 669 LNRNSDKPKKYVSLEGIEYVDRLIDIDQSPIGRTPRSNPATYTGVFDDIRDLFAQTNEAK 728
 L++ KPG + ++G++++D++IDIDQ+PIGRTPRSNPATYTGVFDDIRD+FAQTNEAK
 Sbjct: 661 LHKAKAKPGSHKEIKGLDHLDKVIDIDQAPIGRTPRSNPATYTGVFDDIRDVFAQTNEAK 720

55 Query: 729 IRGYKGRFSFNVKGRCECSGDIKIEMHFLPDVYVPCEVCHGTRYNSETLEVHYKE 788
 +RGYKGRFSFNVKGRCE+C GDGIKIEMHFLPDVYVPCEVCHG RYN ETLEV YK
 Sbjct: 721 VRGYKGRFSFNVKGRCEACRGDGIKIEMHFLPDVYVPCEVCHGKRYNRETLEVITYK 780

60 Query: 789 KNIAQILDMTVNDAVTFFAAIPKIARKLQTIKDVGLGYVTLGQPATTLSCGEAQRMKLAS 848
 K+I+ +LDMTV DA++FF IPKI RKLQT+ DVGLGY+TLGQPATTLSCGEAQR+KLAS
 Sbjct: 781 KSIDVLDMTVEDALSFFENIPKIKRKLQTLVDVGLGYITLGQPATTLSCGEAQRVKLAS 840

Query: 849 ELHKRSTGKSLYLDEPTTGLHADDIARLLKVLDRFVDDGNTVLVIEHNLDVIKTADHII 908
 ELHKRSTG++LYILDEPTTGLH DDIA RL VL R VD+G+TVLVIEHNLD+IKTAD+I+
 Sbjct: 841 ELHKRSTGRTLYILDEPTTGLHVDDIARLLVVLQRLVDNGDNTVLVIEHNLDI IKTADYIV 900

Query: 909 DLGPEGGIGGGQIVAIGTPPEEVAENPKSYTGYYLKEKLAR 948
 DLGPEGG GGG IVA GTPEE+ E +SYTG YLK + R
 Sbjct: 901 DLGPEGGAGGCTIVASGTPPEITEVEESYTGYYLKEKLAR 940

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3559> which encodes the amino acid sequence <SEQ ID 3560>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1285-

bacterial cytoplasm --- Certainty=0.1138(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 835/940 (88%), Positives = 896/940 (94%)

Query: 7 MQDKLMIRGARAHNLKNIISVDIPRDKLVVVTGLSGSGKSSLAFTDIYAEGQRRYVESLSA 66
 MQ+K++I GARAHNLKNI V+IPRDKLVVVTGLSGSGKSSLAFTDIYAEGQRRYVESLSA
 10 Sbjct: 11 MQNKIIIHGARAHNLKNIDVEIPRDKLVVVTGLSGSGKSSLAFTDIYAEGQRRYVESLSA 70

Query: 67 YARQFLGNMEKPDVDSIDGLSPAISIDQKTTSKNPRSTVGTVTTEINDYLRLLYARVGTPY 126
 YARQFLGNMEKPDVDSIDGLSPAISIDQKTTSKNPRSTVGTVTTEINDYLRLLYARVGTPY
 15 Sbjct: 71 YARQFLGNMEKPDVDSIDGLSPAISIDQKTTSKNPRSTVGTVTTEINDYLRLLYARVGTPY 130

Query: 127 CINGHGAITASSVEQIVDKVLALPERTKMQILAPIIRKKGQHKSTFEKIQKDG YVRVRI 186
 CINGHGAITASS EQIV++VLALPERT+MQILAP++RRKKGQHK+ FEKIQKDG YVRVRI+
 Sbjct: 131 CINGHGAITASSAEQIVEQVLALPERTRMQILAFVVRKKGQHKTVFEKIQKDG YVRVRV 190

Query: 187 DGDIDHVTVEPELSKSKMHNIDIVDRLINKEGIRSRLFDSVEAALRLSDGYVVIDTMDG 246
 DGDIDVTVEPELSKSKMHNI++V+DRL+NK+GIRSRLFDSVEAALRL DGY++IDTMDG
 20 Sbjct: 191 DGDIDVTVEPELSKSKMHNIEVVIDRLVNKDGIRSRLFDSVEAALRLDGYLMIDTMDG 250

Query: 247 NELLFSEHYSCPBCGFTVPELEPRLFSFNAPFGSCPTCDGLGIKLEVDIDLVPDRSKTL 306
 NELLFSEHYSCP CGFTVPELEPRLFSFNAPFGSCPTCDGLGIKLEVD+DLV+PD SK+L
 25 Sbjct: 251 NELLFSEHYSCPVCFTVPELEPRLFSFNAPFGSCPTCDGLGIKLEVDLVLVDPDSKSL 310

Query: 307 REGALVPWNPISSNYYPTMLEQAMTQFGVMDTTFPEKLKAEQDLALYSGGEREFHFHYI 366
 REGAL PWNPISSNYYPTMLEQAM FGVDMDTTFE L++ E+DL LYSGG+REFHFHY+
 30 Sbjct: 311 REGALAPWNPISSNYYPTMLEQAMASFGVMDTTFEALTEERDLVLYSGGDEREFHFHYV 370

Query: 367 NDFGGERNIDLPFEGVNNINRRYHETNSDYTRNVMREYMNELKCNTCHGYRLNDQALCV 426
 NDFGGERNID+PFEGVV N+NRRYHETNSDYTRNVMR YMNEL C TCHGYRLNDQALCV
 35 Sbjct: 371 NDFGGERNIDIPFEGVTVNRRYHETNSDYTRNVMRGYMNELTCATCHGYRLNDQALCV 430

Query: 427 RVGGEGLNIGQVSDLSIADHLELLETLRLSSNEQLIARPIIKEIHDRLSFLNNVGLNYL 486
 VGGEEG +IGQ+S+LSIADHL+LLE L L+ NE IA+PI+KEIHDRL+FLNNVGLNYL
 40 Sbjct: 431 HVGGEETHIGQISELSIADHLQLELELTENESTIAKPIVKEIHDRLTLFLNNVGLNYL 490

Query: 487 NLSRSAGTSLGGESQRIRLATQIGSNLSGVLYVLDEPSIGLHQRDNDRLIDSLKKMRDLG 546
 LSR+AGTSLGGESQRIRLATQIGSNLSGVLY+LDEPSIGLHQRDNDRLI+SLKKMRDLG
 45 Sbjct: 491 TLSRAAGTSLGGESQRIRLATQIGSNLSGVLYILDEPSIGLHQRDNDRLIESLKKMRDLG 550

Query: 547 NTLIVVEHDEDTMMAADWLIDVGPAGAFGGEIVASGTPKQVAKNKSITGQYLSGKKVI 606
 NTLIVVEHDEDTMM ADWLIDVGPAG FGGEI ASGTPKQVAKN KSITGQYLSGKK I
 50 Sbjct: 551 NTLIVVEHDEDTMMQADWLIDVGPAGAFGGEITASGTPKQVAKNKSITGQYLSGKKFI 610

Query: 607 PVPSERRVGNRFLIEIKGAAENNLQNLQDVKFLPKFIAVTGVSGSGKSTLINSILKKAVA 666
 PVP ERR GNGRF+EIKGAA+NNLQ+LDV+FPLGKFIATGVSGSGKSTL+NSILKKAVA
 55 Sbjct: 611 PVPLERRSGNRFIEIKGAAQNNLQSLDVRFPPLGKFIATGVSGSGKSTLVNSILKKAVA 670

Query: 667 QKLNARN+DKPGKY S+ GIE+++RLIDIDQSPIGRTPRSNPATYTG VFDDIRDLFAQTNE 726
 QKLNARN+DKPGKY S+ GIE+++RLIDIDQSPIGRTPRSNPATYTG VFDDIRDLFAQTNE
 60 Sbjct: 671 QKLNARNADKPGKYHSISGIEHIERLIDIDQSPIGRTPRSNPATYTG VFDDIRDLFAQTNE 730

Query: 727 AKIRGYKKGRFSFNVKGRCECSGDGIIKIEMHFLPDVYVPCEVCHGRYNSSETLEVHY 786
 AKIRGYKKGRFSFNVKGRCE+CSGDGIIKIEMHFLPDVYVPCEVCHG RYNSSETLEVHY
 65 Sbjct: 731 AKIRGYKKGRFSFNVKGRCEACSGDGIIKIEMHFLPDVYVPCEVCHGRYNSSETLEVHY 790

Query: 787 KEKNIAQILDMTVNDVATFFFAIPKIARKLQTIKDVGLGYVTLGQPATTLSGGEAQRMKL 846
 K KNIA++LDMTV+DA+ FF+AIPKIARK+QTIKDVGLGYVTLGQPATTLSGGEAQRMKL
 70 Sbjct: 791 KGKNIAEVLDMTVDDALVFFSAIPKIARKIQTIKDVGLGYVTLGQPATTLSGGEAQRMKL 850

Query: 847 ASELHKRSTGKSLYILDEPTTGLHADDIARLLKVLDRFVDDGNTVLVIEHNLDVIKTADH 906
 ASELHKRSTGKSLYILDEPTTGLH DDIA RLKVL+RFVDDGNTVLVIEHNLDVIK+ADH
 75 Sbjct: 851 ASELHKRSTGKSLYILDEPTTGLHTDDIARLLKVLDRFVDDGNTVLVIEHNLDVIKSADH 910

-1286-

Query: 907 IIDLGPEGGIGGGQIVAIGTPEEVAENPKSYTGYYLKEKL 946
 IIDLGPEGG GGGQIVA GTPEEVA+ +SYTG+YLK KL
 Sbjct: 911 IIDLGPEGGDGGGQIVATGTPEEVAQVKESYTGHYLKVKL 950

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1147

A DNA sequence (GBSx1223) was identified in *S.agalactiae* <SEQ ID 3561> which encodes the amino acid sequence <SEQ ID 3562>. Analysis of this protein sequence reveals the following:

10 Possible site: 60
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -10.40	Transmembrane	471 - 487 (463 - 490)
INTEGRAL	Likelihood = -9.29	Transmembrane	246 - 262 (242 - 264)
INTEGRAL	Likelihood = -7.27	Transmembrane	183 - 199 (178 - 207)
15 INTEGRAL	Likelihood = -5.41	Transmembrane	351 - 367 (349 - 370)
INTEGRAL	Likelihood = -4.41	Transmembrane	87 - 103 (83 - 107)
INTEGRAL	Likelihood = -3.24	Transmembrane	375 - 391 (374 - 392)
INTEGRAL	Likelihood = -2.97	Transmembrane	17 - 33 (16 - 35)
INTEGRAL	Likelihood = -2.28	Transmembrane	420 - 436 (420 - 438)
20 INTEGRAL	Likelihood = -1.97	Transmembrane	320 - 336 (320 - 337)
INTEGRAL	Likelihood = -1.75	Transmembrane	214 - 230 (214 - 230)
INTEGRAL	Likelihood = -1.75	Transmembrane	288 - 304 (288 - 304)
INTEGRAL	Likelihood = -1.70	Transmembrane	110 - 126 (110 - 126)
25 INTEGRAL	Likelihood = -0.69	Transmembrane	152 - 168 (151 - 168)

----- Final Results -----

bacterial membrane	---	Certainty=0.5161(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12192 GB:Z99106 similar to multidrug resistance protein [Bacillus subtilis]
 Identities = 198/481 (41%), Positives = 300/481 (62%), Gaps = 24/481 (4%)

35 Query: 9 IHGKPYNRTAMITLLLIATFAGVLNQTSLGTAIPTLMNSFNISLSTAQQATTWFLLANGI 68
 I KP+NR+ ++ +LL F +LNQT L TA+P +M FN+ + AQ TT F+L NGI
 Sbjct: 5 IEQKPFNRSVIVGILLAGAFVAILNQTLLITALEPHIMRDFNVDAQWLTTSFMLTNGI 64

40 Query: 69 MIPVSAYLATRFSTKWLYVTSYVLLIGLLMTTLAPTSNWNLFVGRRIQAIISVGISMPL 128
 +IP++A+L +F+++ L +T+ + G ++ AP N+ + L RIIQA GI MPL
 Sbjct: 65 LIPITAFLEIKFTSRALLITAMSIFTAGTVVGAFAF--NFPVLLTARIQAAGAGIMMPL 122

45 Query: 129 MQVVMNVNVPPEQRGAAMGLNGLVVGLAPGAIGPTLAGWILKQEFHFAGHDLTWRAIFLLP 188
 MQ V + +FP E+RG AMG+ GLV+ APAIGPTL+GW ++ +WR++F +
 Sbjct: 123 MQTVFLTIFPIEKRGQAMGMVGLVISFAPAIGPTLSGWAVEA-----FSWRSIFYII 174

Query: 189 LLILTVTITLSPFVLKDVVDNKS VKLEVPSLILSIIGFGSFLWGFTNVATYWGWDIGYVI 248

-1287-

Query: 429 LSSVAQNIITNNKPSKDLLTMNPLKYANQMLNASLDGFHVSFAIGFVFAVLGLLVSLFLRK 489
 L SV N + + +A+L G + +F + V A++G L+S L+K
 Sbjct: 414 LVSVMNQAAH-----AGTTNVKHAALHGMNAAFIVAAVIALVGFLLSFTLKK 461

5 There is also homology to SEQ ID 46.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1148

10 A DNA sequence (GBSx1224) was identified in *S.agalactiae* <SEQ ID 3563> which encodes the amino acid sequence <SEQ ID 3564>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -8.81 Transmembrane 8 - 24 (5 - 30)
 INTEGRAL Likelihood = -7.32 Transmembrane 36 - 52 (31 - 54)
 15 ----- Final Results -----
 bacterial membrane --- Certainty=0.4524(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 20

A related GBS nucleic acid sequence <SEQ ID 10109> which encodes amino acid sequence <SEQ ID 10110> was also identified.

A related GBS gene <SEQ ID 8743> and protein <SEQ ID 8744> were also identified. Analysis of this protein sequence reveals the following:

25 Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: 9.52
 GvH: Signal Score (-7.5): -3.4
 Possible site: 22
 >>> Seems to have an uncleavable N-term signal seq
 30 ALOM program count: 1 value: -7.32 threshold: 0.0
 INTEGRAL Likelihood = -7.32 Transmembrane 11 - 27 (6 - 29)
 PERIPHERAL Likelihood = 11.19 130
 modified ALOM score: 1.96
 35 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 SEQ ID 8744 (GBS29) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 2; MW 25.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 6; MW 51kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1288-

Example 1149

A DNA sequence (GBSx1225) was identified in *S.agalactiae* <SEQ ID 3565> which encodes the amino acid sequence <SEQ ID 3566>. This protein is predicted to be aminopeptidase P (pepQ). Analysis of this protein sequence reveals the following:

```

5   Possible site: 41
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.0724(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAA70068 GB:Y08842 aminopeptidase P [Lactococcus lactis]
    Identities = 44/126 (34%), Positives = 78/126 (60%)

    Query: 6   RLTRCQTAISQLSCDALLITNLTNIFYLTGFGSGTNATVLISPKHRIFVTDSRYALIAKNT 65
              R+ + + + + D+LLIT++ NIFYLTGFGSGT TV ++ K IF+TDSRY+ +A+
    Sbjct: 2   RIEKLKVKMLTENIDSLITDMKNIFYLTGFGSGTAGTVFLTQKRNI FMTDSRYSEMARGL 61

20   Query: 66  VREFDIISREPLAAILKIITRDDALIAIGFETDISYHMYKHMVEVFEDYRLIEAPSVVEK 125
              ++ F+II +R+P++ + ++ +++ + FE + Y +K + + L + V +
    Sbjct: 62  IKNFEIIEITRDPISLLTELSASESVKNMAFEETVDYAFFKRLSKAATKLDLFSTSNFVLE 121

25   Query: 126 LRMIKD 131
              LR IKD
    Sbjct: 122 LRQIKD 127

```

There is also homology to SEQ ID 3568.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1150

A DNA sequence (GBSx1226) was identified in *S.agalactiae* <SEQ ID 3569> which encodes the amino acid sequence <SEQ ID 3570>. This protein is predicted to be aminopeptidase P (pepQ-2). Analysis of this protein sequence reveals the following:

```

35   Possible site: 44
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40          bacterial cytoplasm --- Certainty=0.2508(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

45   >GP:CAA70068 GB:Y08842 aminopeptidase P [Lactococcus lactis]
    Identities = 131/205 (63%), Positives = 163/205 (78%), Gaps = 3/205 (1%)

    Query: 2   LDFIKPDRITELQVANFLDFRMRELGATGPFSDFIVASGYRSAMPHGVASQKTIQSGETL 61
              L FI+P RT E++VANFLDF+MR+L A+G SF+ IVASG RS++PHGVA+ K IQ G+ +
50   Sbjct: 149 LRFIEBPGRT-EIEVANFLDFKMRDLEASGISFETIVASGKRSSLPHGVATSKMLIQFGDPV 207

    Query: 62  TLDFGCYQHYVSDMTRTIHIGHVTDQEREIYDIVLKSNAIIGNVKSGMKRCDYDYLAR 121
              T+DFGCYY+HY SDMTRTI +G V D+ R IY+ V K+N+A+I VK+GM YD + R
    Sbjct: 208 TIDFGCYEYHVASDMTRTIFVGSVDKMRTIYETVRKANEALIKQVKAGMTYAQYDNIPR 267

55   Query: 122 QVIENSGYGNHFTHGIGHGMGLDVHEIPYFGKS--EGVIASGMVVTDPEGIYLDNKYGV 179

```

-1289-

+VIE + +G +FTHGIGHG+GLDVHEIPYF +S E + SGMV+TDEPGIYL GVR
 Sbjct: 268 EVIEKADFGQYFTHGIGHGLGLDVHEIPYFNQSMTENQLRSGMVITDEPGIYLPFEGGV 327

5 Query: 180 IEDDLLITETGCEVLTSAPKELIVL 204
 IEDDLL+TE GCEVLTSAPKELIV+
 Sbjct: 328 IEDDLLVTENGCEVLTKAPKELIVI 352

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3567> which encodes the amino acid sequence <SEQ ID 3568>. Analysis of this protein sequence reveals the following:

10 Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 15 bacterial cytoplasm --- Certainty=0.1450(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 145/203 (71%), Positives = 171/203 (83%)

20 Query: 2 LDFIKPDRTELQVANFLDFRMRELGATGPSFDFIVASGYRSAMPHGVASQKTIQSGETL 61
 LDFIKP TTE +ANFLDFRMR+ GA+G SFD IVASGY SAMPHG AS K IQ+ E+L
 Sbjct: 168 LDFIKPGTTTERDLANFLDFRMROYGASGTSFDIIVASGYLSAMPHGRASDKVIQNKESL 227

25 Query: 62 TLDFGCYYQHYSMDMTRTIHIGHVTDQEREIYDIVLKSQAIIGNVKSGMKRCDYDYLAR 121
 T+DFGCYY HYVSDMTRTIHIG VTD+EREIY +VL +N+A+I +GM D+D + R
 Sbjct: 228 TMDFGCYNHYSMDMTRTIHIGQVTDEEREIYALVLAANKALIAKASAGMTYSDFDGIPIR 287

30 Query: 122 QVIENSGYGNHFTHGIGHGMLDVHEIPYFGKSEGVASGMVVTDEPGIYLDNKYGVRIE 181
 Q+I +GYG+ FTHGIGHG+GLD+HE P+FGKSE ++ +GMVVTDEPGIYLDNKYGVRIE
 Sbjct: 288 QLITEAGYGSRFTHGIGHGIGLDIHENPFFGKSEQLLQAGMVVTDEPGIYLDNKYGVRIE 347

Query: 182 DDLITETGCEVLTSAPKELIVL 204
 DDL+IT+TGC+VLT APKELIVL
 35 Sbjct: 348 DDLVITKTGCQVLTAPKELIVL 370

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1151

40 A DNA sequence (GBSx1227) was identified in *S.galactiae* <SEQ ID 3571> which encodes the amino acid sequence <SEQ ID 3572>. This protein is predicted to be yfhC protein (comEB). Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1401(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05053 GB:AP001511 late competence operon required for DNA
 binding and uptake [Bacillus halodurans]
 Identities = 78/146 (53%), Positives = 107/146 (72%)

55 Query: 1 MNRLSWEDYFMANAELISKRSTCDRAVFGAVLVKNNRIIATGYNGGVSETDNCNEVGHY 60
 MNR+SW+ YFMA + L++ RSTC R VGA +V++ RIIA GYNG +S +C + G Y+
 Sbjct: 1 MNRISWDQYFMAQSHLLALRSTCTRLMVGATIVRDKRIIAGGYNGSISGGPHCIDEGCYV 60

-1290-

Query: 61 EDGHCIRTVHAEMNALIQCAKEGISTNNTETIYVTHFPCINCTKALLQAGVKKITYKANYR 120
 +GHCIRT+HAE+NAL+QCAK G+ T EIYVTHFPC+NCTKA++Q+G+KK+ Y +Y+
 Sbjct: 61 VEGHCIRTIHAEVNALLQCAKFGVPTGAEIYVTHFPCVNCTKAI IQSGIKKVVYATDYK 120

5

Query: 121 PHPPAIELMEAKGVAYVQHDVPEVTL 146
 P+A EL GV Q ++ E+ L
 Sbjct: 121 NSPYAEELFRDAGVDVEQVELEEMIL 146

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3573> which encodes the amino acid sequence <SEQ ID 3574>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3155(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 133/146 (91%), Positives = 140/146 (95%)

- Query: 2 NRLSWEDYFMANAELISKRSTCDRAFGAVLVKNNRIIATGYNGGVSETDNCNEVGHYME 61
 NRLSW+DYFMANAELISKRSTCDRAFGAVLVK+NR I IATGYNGGV S TDNCNE GHYME
- 25 Sbjct: 18 NRLSWQDYFMANAELISKRSTCDRAFGAVLVKDNRIIATGYNGGV SATDNCNEAGHYME 77
- Query: 62 DGHHCIRTVHAEMNALIQCAKEGISTNNTETIYVTHFPCINCTKALLQAGVKKITYKANYRP 121
 DGHHCIRTVHAEMNALIQCAKEGIST+ TETIYVTHFPCINCTKALLQAG+ KITYKA+YRP
- 30 Sbjct: 78 DGHHCIRTVHAEMNALIQCAKEGISTDGTETIYVTHFPCINCTKALLQAGITKITYKAHYRP 137
- Query: 122 HPFAIELMEAKGVAYVQHDVPEVTLG 147
 HPFAIELME KGVAYVQHDV++ LG
- Sbjct: 138 HPFAIELMEKKG VAYVQHDV PQIVLG 163

- 35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1152

A DNA sequence (GBSx1228) was identified in *S.agalactiae* <SEQ ID 3575> which encodes the amino acid sequence <SEQ ID 3576>. Analysis of this protein sequence reveals the following:

- 40 Possible site: 13
 >>> Seems to have no N-terminal signal sequence
- Final Results -----
 bacterial cytoplasm --- Certainty=0.2454(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
- 45

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1291-

Example 1153

A DNA sequence (GBSx1229) was identified in *S.agalactiae* <SEQ ID 3577> which encodes the amino acid sequence <SEQ ID 3578>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 25
      >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -1.65    Transmembrane    4 - 20 ( 3 - 21)

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1154

A DNA sequence (GBSx1230) was identified in *S.agalactiae* <SEQ ID 3579> which encodes the amino acid sequence <SEQ ID 3580>. Analysis of this protein sequence reveals the following:

```

20      Possible site: 54
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
25      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

30      >GP:BA04699 GB:AP001510 unknown conserved protein [Bacillus halodurans]
      Identities = 47/94 (50%), Positives = 65/94 (69%)

      Query: 2  LLPVGSVVYLIDGNQKLVIVNRGAIVEQEGQEVYFDYLGIGIFPEGLNLEQVYYFNQEDID 61
      +LP+GS+VYL+G KL+I+NRG I+E G+ FDY G +P+GL ++V+YFN E+ID
35      Sbjct: 1  MLPIGSIVYLKEGTSKLMILNRGPILLEANENKMFYSGCFYPQGLVDPKVFYFNHENID 60

      Query: 62  EVVFEGYHDEEEERVSRLLIEKWKNTEGKNLPK GK 95
      EVVFEG+ D+EE+R +L WK KGK
      Sbjct: 61  EVVFEGFQDDEEQRFQKLFHDWKKENKDRYVKGK 94

```

40 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1155

45 A DNA sequence (GBSx1231) was identified in *S.agalactiae* <SEQ ID 3581> which encodes the amino acid sequence <SEQ ID 3582>. Analysis of this protein sequence reveals the following:

```

      Possible site: 15
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.3560(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-1292-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1156

- 10 A DNA sequence (GBSx1232) was identified in *S.agalactiae* <SEQ ID 3583> which encodes the amino acid sequence <SEQ ID 3584>. This protein is predicted to be elongation factor p (efp). Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 15 bacterial cytoplasm --- Certainty=0.3067(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 20 >GP:CAB14376 GB:Z99116 elongation factor P [Bacillus subtilis]
 Identities = 89/186 (47%), Positives = 120/186 (63%), Gaps = 1/186 (0%)
- Query: 1 MIEASKLKAGMTFETADGKLIRVLEASHHKPGKGNTIMRMKLRDVRTGSTFDTSYRPEEK 60
 MI + + G+T + DG + RV++ H KPGKG +R KLR++RTG+ + ++R EK
 25 Sbjct: 1 MISVNDPRTGLTIDV-DGGIWRVVDVFQHVKPGKGAFFVRSKLRNLRTGAIQKTFRAGEK 59
- Query: 61 FEQAI IETVPAQVLYKMDDTAYFMNNETYDQYEIPTVNIENELLYILENSEVVKIQFYGTE 120
 +A IET QVLY D FM+ +Y+Q E+ IE EL Y+LEN V I Y E
 30 Sbjct: 60 VAKAQIETKTMQVLYANGDQHFVMDTSSYEQLELSATQIEEELKYLLNMSVHIMMYQDE 119
- Query: 121 VIGVQIPTITVELTVAETQPSIKGATVTGSGKPAIMETGLVNVNPDFIEAGOKLVINTAEG 180
 +G+++P TVEL V ET+P IKG T +G KPA ETGLVNVNPF+ G LV+NT++G
 35 Sbjct: 120 TLGIELPNTVELKVVETEPGIGKDTASGGTKPAKTETGLVNVNPFVNEGDTLVVNTSDG 179
- Query: 181 TYVSRA 186
 +YVSRA
 40 Sbjct: 180 SYVSRA 185

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3585> which encodes the amino acid
 40 sequence <SEQ ID 3586>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 45 bacterial cytoplasm --- Certainty=0.1813(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 50 Identities = 170/186 (91%), Positives = 180/186 (96%), Gaps = 1/186 (0%)
- Query: 1 MIEASKLKAGMTFETADGKLIRVLEASHHKPGKGNTIMRMKLRDVRTGSTFDTSYRPEEK 60
 MIEASKLKAGMTFE A+GKLIRVLEASHHKPGKGNTIMRMKLRDVRTGSTFDT+YRP+EK
 55 Sbjct: 1 MIEASKLKAGMTFE-AEGKLIRVLEASHHKPGKGNTIMRMKLRDVRTGSTFDTTYRPDEK 59
- Query: 61 FEQAI IETVPAQVLYKMDDTAYFMNNETYDQYEIPTVNIENELLYILENSEVVKIQFYGTE 120

-1293-

FEQAIITVPAQYLYKMDDTAYFMN +TYDQYEIP N+E ELLYILENS+VKIQFYG+E
 Sbjct: 60 FEQAIITVPAQYLYKMDDTAYFMNTDYDQYEIPVANVEQELLYILENSDVKIQFYGSE 119
 Query: 121 VIGVQIPTTVELTVAETQPSIKGATVTGSGKPAIMETGLVNVNVPDFIEAGQKLVINTAEG 180
 5 VIGV +PTTVELTVAETQPSIKGATVTGSGKPAI+ETGLVNVNVPDFIEAGQKL+INTAEG
 Sbjct: 120 VIGVTVPPTTVELTVAETQPSIKGATVTGSGKPAIETGLVNVNVPDFIEAGQKLIINTAEG 179
 Query: 181 TYVSRA 186
 TYVSRA
 10 Sbjct: 180 TYVSRA 185

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1157

- 15 A DNA sequence (GBSx1233) was identified in *S.agalactiae* <SEQ ID 3587> which encodes the amino acid sequence <SEQ ID 3588>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1508(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06505 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 42/107 (39%), Positives = 70/107 (65%), Gaps = 4/107 (3%)

30 Query: 5 NLGEIVISPRVLEVITGIAATKVDGVHSLRNK---AVTDSLKSKSLGRGVYLNKNEEDDTV 61
 +LG + ISP V+EVI GIAA++V+GV ++R V + L K+ G+GV + + D+ +
 Sbjct: 15 DLGRVEISPEVIEVIAGIAASEVEGVATMRGNFAAGVAEKLGYKNHKGKGVKV-DLNDEGI 73

Query: 62 AADIYVYLQYGVNVPVAVSIAIQQAVKTAVYDMAEVKISSVNIHVEGI 108
 D+ V + YGV+VP V+ IQQ +K A+ M +++ S+N+H+ G+

35 Sbjct: 74 IVDVSVIILYGVSVPEVAKKIQQNIQALQMTAIELQSINVHIVGV 120

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3589> which encodes the amino acid sequence <SEQ ID 3590>. Analysis of this protein sequence reveals the following:

40 Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0882(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 101/129 (78%), Positives = 113/129 (87%)

50 Query: 1 MTTENLGEIVISPRVLEVITGIAATKVDGVHSLRNKAVTDSLKSKSLGRGVYLNKNEEDDT 60
 MTTE +GEIVISPRVLEVITGIA T+V+GVHSL NK + DS +K SLG+GVYL+ EED +
 Sbjct: 1 MTTEYIGEIVISPRVLEVITGIATTQVEGVHSLHNKKMADSFNKASLGKGVYLTQTEEDGS 60
 Query: 61 VAADIYVYLQYGVNVPVAVSIAIQQAVKTAVYDMAEVKISSVNIHVEGIVPEKTPKPDLLS 120
 55 V ADIYVYLQYGV VP VS+ IQ+ VK+AVYDMAEV IS+VNIHVEGIV EKTPKPDLLS
 Sbjct: 61 VTADIYVYLQYGVKVPVSMNIQKTVKSAVYDMAEVPISAVNIHVEGIVAEKTPKPDLLS 120
 Query: 121 LFDEDFLDD 129
 LFDEDFLDD

-1294-

Sbjct: 121 LFDEDFLDD 129

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1158

A DNA sequence (GBSx1234) was identified in *S.agalactiae* <SEQ ID 3591> which encodes the amino acid sequence <SEQ ID 3592>. This protein is predicted to be n utilization substance protein b homolog (nusB). Analysis of this protein sequence reveals the following:

Possible site: 27
 10 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.32 Transmembrane 48 - 64 (47 - 64)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
 15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14363 GB:Z99116 similar to transcription termination
 20 [Bacillus subtilis]
 Identities = 51/129 (39%), Positives = 82/129 (63%), Gaps = 9/129 (6%)
 Query: 9 RRDLRERAFQTLFSLETGGEFIDAAHFAYGYDKTVSEDKVLEVPFLLNLVNGVVDHKDE 68
 RR RE+A Q LF ++ ++ A + + E+K F LV+GV++H+D+
 25 Sbjct: 3 RRTAREKALQALFQIDVSDIAVNEA-----IEHALDEEKT---DPFFEQLVHGVLEHQDQ 54
 Query: 69 LDTLISSHLKSGWSLERLTLVDKSLRLGLYEIKYFDETPDRVALNEIIEIAKKYSDETS 128
 LD +IS HL + W L+R+ VD+++LRL YE+ Y ++ P V++NE IE+AK++ D+ +
 30 Sbjct: 55 LDEMISKHLVN-WKLDRIANVDRAILRLAAYEMAYAEDIPVNVSMNEAIELAKRFGDDKA 113
 Query: 129 AKFVNGLLS 137
 KFVNG+LS
 Sbjct: 114 TKFVNGVLS 122

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3593> which encodes the amino acid sequence <SEQ ID 3594>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 40 INTEGRAL Likelihood = -1.75 Transmembrane 53 - 69 (53 - 69)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14363 GB:Z99116 similar to transcription termination
 [Bacillus subtilis]
 Identities = 47/134 (35%), Positives = 76/134 (56%), Gaps = 10/134 (7%)
 50 Query: 15 RRDLRERAFQALFNIEMGAELLAASQFAYGYDKVTGEDAQVLELPFLLSLVTGVNNHKE 74
 RR RE+A QALF I++ +++ + D+ + F LV GV H++
 Sbjct: 3 RRTAREKALQALFQIDV-SDIAVNEAIEHALDEEKTDP-----FFEQLVHGVLEHQD 53
 55 Query: 75 ELDNLISTHLKKGWSLERLTLTDKTLRLGLFEIKYFDKTPDRVALNEIIEVVKYSDET 134
 +LD +IS HL W L+R+ D+ +LRL +E+ Y + P V++NE IE+ K++ D+
 Sbjct: 54 QLEDEMISKHLVN-WKLDRIANVDRAILRLAAYEMAYAEDIPVNVSMNEAIELAKRFGDDK 112

-1295-

Query: 135 SAKFINGLLSQYVS 148
 + KF+NG+LS S
 Sbjct: 113 ATKFVNGVLSNIKS 126

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 104/142 (73%), Positives = 125/142 (87%), Gaps = 1/142 (0%)

Query: 1 MTSVFKDSRRDLRERAFQTLFSLETGGEFIDAHAHFAYGYDKTVSED-KVLEVPIFLLNLV 59
 MT+ F++SRRDLRERAFQ LF++E G E + A+ FAYGYDK ED +VLE+PIFLL+LV
 10 Sbjct: 7 MTNSFQNSRRDLRERAFQALFNIEGMAELLAASQFAYGYDKVTGEDAQVLELPIFLLSLV 66
 Query: 60 NGVVDHKDELDTLISHLKSGWSLERLTLVDKSLRLGLYEIKYFDETPDRVALNEIIEI 119
 GV +HK+ELD LIS+HLK GWSLERLTL DK+LLRLGL+EIKYFD+TPDRVALNEIIE+
 15 Sbjct: 67 TGVNNHKEELDNLISTHLKKGWSLERLTLTDKTLRLGLFEIKYFDKTPDRVALNEIIEV 126
 Query: 120 AKKYSDETSAKFVNGLLSQFIT 141
 KKYSDETSAKF+NGLLSQ+++
 Sbjct: 127 VKKYSDETSAKFINGLLSQYVS 148

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1159

A DNA sequence (GBSx1235) was identified in *S.agalactiae* <SEQ ID 3595> which encodes the amino acid sequence <SEQ ID 3596>. Analysis of this protein sequence reveals the following:

25 Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -2.81 Transmembrane 239 - 255 (239 - 255)
 30 ----- Final Results -----
 bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAC31628 GB:U46902 ScrR [Streptococcus mutans]
 Identities = 225/320 (70%), Positives = 273/320 (85%)
 Query: 1 MVAKLTDVAALAGVSPPTTVSRVINKKGYLSQKTVTKVNEAMRTLGYKPNNLARSLQGKSA 60
 MVAKLTDVA LAGVSPPTTVSRVIN+KGYS+KT+TKV AM+TLGYKPNNLARSLQGKSA
 40 Sbjct: 1 MVAKLTDVAKLAGVSPPTTVSRVINRKGYLSEKTTTKVQAAMKTLGYKPNNLARSLQGKSA 60
 Query: 61 KLIGLIFPNIRNIFYAELIEHLEIELFKHGYKTILCNSEKDPIKEKEYLEMLGANQVDGI 120
 KLIGLIFPNI +IFY+ELIE+LEIELFKHGYK I+CNS+ +P KE++YLEML ANQVDGI
 45 Sbjct: 61 KLIGLIFPNISHIFYSELIEYLEIELFKHGYKAIICNSQNNPDKERDYLEMLEANQVDGI 120
 Query: 121 ISSSHNLGIDDYKVEAPIVAFDRNLAPHIPVSSDNFFGGKMAAQTLLKKHGCQKMIMIT 180
 ISSSHNLGIDDYKVI API+AFDRNLAP+IPVSSDNF GG+MAA+ LKKHGCQ IMI
 Sbjct: 121 ISSSHNLGIDDYKVSAPIIAFDRNLAPNIPVSSDNFEGGRMAAKLLKKHGCQHPIMIA 180
 50 Query: 181 GNDNSDSPTGLRRLGFSYESKESKITVTINGLSNMRREMEKSIISTHKPDGIFTSDDLT 240
 G DNS+SPT LR+LGF ++ + ++ LS +R+EME+K I+ KPDGIF SDD+T
 Sbjct: 181 GKDNNSPTALRQLGFKSVFAQAPIFHLSELSITRKEMEIKVILQNEKPDGIFLSDDMT 240
 Query: 241 ALLVIKLISQLGLSIPEDIKVIYDGTSTFIQDYVPHLTITIKQFIREIAQLMVEILLAKIE 300
 A+L +K+ +QL ++IP ++K+IGYDGT F+++Y P+LTTI+QPI++IA L+V+ILL KI+
 55 Sbjct: 241 AILTMKIANQLNITIPHEKLIIGYDGTDFVENYYPYLTITIQPIKDIAHLLVDILLKKID 300
 Query: 301 GQKTNKDYILPVSLIPGSSV 320
 Q KDYILPV L+ G SV
 60 Sbjct: 301 HQDIPKDYILPVGLLSGESV 320

-1296-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3597> which encodes the amino acid sequence <SEQ ID 3598>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC31628 GB:U46902 ScrR [Streptococcus mutans]

Identities = 226/321 (70%), Positives = 269/321 (83%), Gaps = 1/321 (0%)

Query: 1 VVAKLTDVAALAGVSPPTTVSRVINKKGYLSQKTVNKNKAMRELGYKPNNLARSLSQKST 60
+VAKLTDVA LAGVSPPTTVSRVIN+KGYLS+KT+ KV AM+ LGYKPNNLARSLSQKGS
Sbjct: 1 MVAKLTDVAKLAGVSPPTTVSRVINRKGYLSEKTTTKVQAAMKTLGYKPNNLARSLSQKSA 60

Query: 61 QLIGLIFPNISNIFYAELIEHLEIELFKQGYKTIICNSEHNPVKEREYLEMLAANQVDGI 120
+LIGLIFPNIS+IFY+ELIE+LEIELFK GYK IICNS++NP KER+YLEML ANQVDGI
Sbjct: 61 KLIGLIFPNISHIFYSELIEYLEIELFKHGYKAIICNSQNNPKERDYLEMLEANQVDGI 120

Query: 121 ISSSHNLGIEDYERVEAPIVAFDRNLAPNIPVISSDNFEGGKLAQAQTLQKHGCQNIVMIT 180
ISSSHNLGI+DYE+V API+AFDRNLAPNIP++SSDNFEGG++AA+ L+KHGCQ+ +MI
Sbjct: 121 ISSSHNLGIDDYKVSAPIIAFDRNLAPNIPVSSDNFEGGRMAAKLLKKHGCQHPIMIA 180

Query: 181 GNDNSDSPTGLRQLGFNYQLKRSAEI IKLPNNLSPVRREMEIKSILATRKPDGLFVSDDL 240
G DNS+SPT LRQLGF + A I L LS +R+EMEIK IL KPDG+F+SDD+
Sbjct: 181 GKDNSNSPTALRQLGFK-SVFAQAPIFHLSGELSIIRKEMEIKVILQNEKPDGIFLSDDM 239

Query: 241 TAILIMKVAQLHITIPEDMKVIGYDGTTFIQYVPQLATIRQPIDETIAKLSVEILIKKI 300
TAIL MK+A QL+ITIP ++K+IGYDGT F++ Y P L TIRQPI +IA L V+IL+KKI
Sbjct: 240 TAILTMKIANQLNITIPHELKIIGYDGTHTFVENYYPYLTIRQPIKDIAHLLVDILKKI 299

Query: 301 KKEKTSKDYILPITLLPGASI 321
+ KDYILP+ LL G S+
Sbjct: 300 DHQDIPKDYILPVGLLSGESV 320

An alignment of the GAS and GBS proteins is shown below.

Identities = 247/321 (76%), Positives = 293/321 (90%), Gaps = 1/321 (0%)

Query: 1 MVAKLTDVAALAGVSPPTTVSRVINKKGYLSQKTVNKNKAMRELGYKPNNLARSLSQKSA 60
+VAKLTDVAALAGVSPPTTVSRVINKKGYLSQKTV KVN+AMR LGYKPNNLARSLSQKGS
Sbjct: 1 VVAKLTDVAALAGVSPPTTVSRVINKKGYLSQKTVNKNKAMRELGYKPNNLARSLSQKST 60

Query: 61 KLIGLIFPNIRNIFYAELIEHLEIELFKHGYKTIICNSEKDPIKEKEYLEMLGANQVDGI 120
+LIGLIFPNI NIFYAELIEHLEIELFK GYKTI+CNSE +P+KE+EYLEML ANQVDGI
Sbjct: 61 QLIGLIFPNISNIFYAELIEHLEIELFKQGYKTIICNSEHNPVKEREYLEMLAANQVDGI 120

Query: 121 ISSSHNLGIDDYKVEAPIVAFDRNLAPHIPVSSDNFEGGKMAAQTLLKKHGCQKMMIT 180
ISSSHNLGI+DYE+VEAPIVAFDRNLAP+IP++SSDNF GSK+AAQTL+KHGCQ ++MIT
Sbjct: 121 ISSSHNLGIEDYERVEAPIVAFDRNLAPNIPVISSDNFEGGKLAQAQTLQKHGCQNIVMIT 180

Query: 181 GNDNSDSPTGLRRLGFSYESKES-KVITVTNGLSNMRREMEIKSIISTHKPDGIFTSDDL 239
GNDNSDSPTGLR+LGF+Y+ K S ++I + N LS +RREME+KSI++T KPDG+F SDDL
Sbjct: 181 GNDNSDSPTGLRQLGFNYQLKRSAEI IKLPNNLSPVRREMEIKSILATRKPDGLFVSDDL 240

Query: 240 TALLVIKLISQLGLSIPEDIKVIYDGTSTFIQDYVPHLTTIKQPIREIAQLMVEILLAKI 299
TA+L++K+ QL ++IPED+KVIYDGT+FIQ YVP L TI+QPI EIA+L VEIL+ KI
Sbjct: 241 TAILIMKVAQLHITIPEDMKVIGYDGTTFIQYVPQLATIRQPIDETIAKLSVEILIKKI 300

Query: 300 EGQKTNKDYILPVSLIPGSSV 320
+ +KT+KDYILP++L+PG+S+
Sbjct: 301 KKEKTSKDYILPITLLPGASI 321

-1297-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1160

- 5 A DNA sequence (GBSx1236) was identified in *S.agalactiae* <SEQ ID 3599> which encodes the amino acid sequence <SEQ ID 3600>. This protein is predicted to be sucrose-6-phosphate hydrolase (cscA). Analysis of this protein sequence reveals the following:

Possible site: 52
>>> Seems to have no N-terminal signal sequence

10

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4775 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

15

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA35872 GB:X51507 sucrose-6-phosphate hydrolase [Streptococcus mutans]
Identities = 303/479 (63%), Positives = 359/479 (74%), Gaps = 25/479 (5%)

20

Query: 1 MNLPTIEIRYPYDEWTEEDKENIVKNVSKSPWRATYHLEAKTGLLNDPNGFSYFNGKFLH 60
MNLPT IRYR Y +WTEE+ ++I NV+ SPW TYH+E KTGLLNDPNGFSYFNGKFL+L
Sbjct: 1 MNLPTQIRYRRYQDWTEEEIKSIKTNVALSPWHTTYHIEPKTGLLNDPNGFSYFNGKFLN 60

25

Query: 61 FYQNWPFGAHGLKQWVHTESDDLVEHFKETGIKLPDHDVNDSHGAYSGSALAIIDDKLFLF 120
FYQNWPFGAHGLK W+HTES+DLVHFKETG L PD +DSHGAYSGSA I D+LFLF
Sbjct: 61 FYQNWPFGAHGLKSWIHTESDDLVEHFKETGTVLYPDTSHDHSHGAYSGSAYEIGDQLFLF 120

30

Query: 121 YTGNVDRMDKWNDRPRQIGAWMTNDGKITKFDKVLISQPNVDTEHFRDPQIFNYDNQFYAV 180
YTGNVDR W R P QIGA+M G I KF VLI QPNVDTEHFRDPQIFNY QFYA+
Sbjct: 121 YTGNVDRDENWVRHPLQIGAFMDKKNIQKFTDVLIKQPNVDTEHFRDPQIFNYKQGFYAI 180

35

Query: 181 IGAQNSKKCGFIKLYKALNNDIHHWEFVGDLDFGGTGSEYMIECPNII FVKGKPVLLYSP 240
+GAQ+ LDFGG+ SEYMIECPN++F+ +PVL+YSP
Sbjct: 181 VQAQS-----LDFGGSKSEYMIECPNLVFINEQPVLIIYSP 215

40

Query: 241 QGLDKNELDYQNIYPNTYKIGQYFDANSSKIVEPSPIYNLDYGFEAYATQGFNTSDGRA 300
QGL K+ELDY NIYPNTYK+ Q FD +V+ S I NLD+GFE YATQ FN DGR +
Sbjct: 216 QGLSKSELDYHNIYPNTYKVCQSFDTEKPAVDASEIQNLDFGFEYATQAFNAPDGRVY 275

45

Query: 361 NNTYELELLVPRNDLSSFVLFANPKGQGLSITIDTVKGKVIIDRSQAGQYATEFGTSRQ 420
NNTYELEL + ++ +LFA+ KG GL+IT+DT G ++IDRS+AG+QYA EFG+ R
Sbjct: 336 NNTYELELTFDSSSVNELLLFADNKGNGLAITVDTKMGITILIDRSKAGEQYALEBFGSQRS 395

50

Query: 421 CDIPKDATSINIFIDKSI FIFINKGEKVFTGRVFPDABQSGIQLKEGHVHGKYFELKY 479
C I T +NIF+DKSIFEIFINKGEKVFTGRVFP+ +Q+GI +K G G Y+ELKY
Sbjct: 396 CSIQAKETVNVNIFVDKSI FIFINKGEKVFTGRVFPNDKQTGIVIKSGKPSGNYELKY 454

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3601> which encodes the amino acid sequence <SEQ ID 3602>. Analysis of this protein sequence reveals the following:

- 55 Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4629 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-1298-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 288/479 (60%), Positives = 367/479 (76%)

5
Query: 1 MNLPTAIRYRPFYDEWTEEDKENIVKNSKSPWRATYHLEAKTGLLNDPNGFSYFNGKFHL 60
M+LP IRYRPFY EW+ +D + I + +++SPW + +H+E KTGLNDPNGFSYFNG++HL
Sbjct: 2 MDLPQAIRYRPFYKEWSSKDYQAITEKMAQSPWHSQPHVEPKTGLLNDPNGFSYFNGRYHL 61

10
Query: 61 FYQNWPFGAHGLKQWVHTESDDLHVHFKETGIKLPDHDVNDSHGAYSGSALAIIDDKLFLF 120
FYQNWPF+GAHGLKQWVH S DLVHF ET +L PDH +DSHGAYSGSA AIDDKLFLF
Sbjct: 62 FYQNWPFYGAHGLKQWVHMTSTDLVHFTETSRLLPDHAHDSHGAYSGSAYAIIDDKLFLF 121

15
Query: 121 YTGNVDRMDKWNDRDPQIGAWMTNDGKITKFDKVLISQPNVDTEHFRDPQIFNYDNQFYAV 180
YTGNVDRD W R P Q+GAWM G I+K +VLI QP+DVTEHFRDPQ+F+Y QFYA+
Sbjct: 122 YTGNVDRDANWVRTPQLVGAWMDKQGNISKIPQVLIEQDDVTEHFRDPQLFSYQGGFYAI 181

20
Query: 181 IGAQNSKKCGFIKLYKALNNDIHHWEFVGDLDGFGTGSEYMECPNII FVKGKPVLLYSP 240
IGAQ G IKLYKA++N + +W F+ DLDF +G+EYMECPN++FV KPVLL++SP
Sbjct: 182 IGAQGLDGKGIKLYKAVDNHDNWRFIADLDFFDSGTEYMECPNLVFDVDDKPVLLIFSP 241

25
Query: 241 QGLDKNELDYQNIYPNTYKIGQYFDANSSKIVEFSPINLDYGFAYATQGFNTSDGRA 300
QGL K +LDYQNIYPNTYKI + F+ + +++ + NLD+GFAYATQ F++ DGR
Sbjct: 242 QGLAKADLDYQNIYPNTYKIFESFPNTPETGQLLGGGALQNLDFGFAYATQAFSSPDGRVL 301

30
Query: 301 IVSWIGLPDIDYPSDQFDYQGAMSLVKELSIKNGNLYQYVPVAMKNLRQHQAEFKTQLQT 360
VSWIGLPDIDYP+D++DYQGA+SLVKEL IK+G LYQ PV A++NLR F ++ +
Sbjct: 302 AVSWIGLPDIDYPTDRDYQGALSIVKELRIKDGILYQTPVSALQNLRGPAELFHNKIDS 361

35
Query: 361 NNTYELELLVPRNDLSSFVLFPANPKGQGLSITIDTVKGVIIIDRSQAGQQYATEFGTSRQ 420
+N YELEL +P +LFA+ KG GL + +DT KG++ IDRS+AG QYA ++GT R
Sbjct: 362 SNCYELELTIPGQKLDLLLFADQKGNLRLKVDTTKQQLSIDRSRAGVQYAQDYGTVRS 421

40
Query: 421 CDIPKDATSINIFIDKSIFEIFINKGEKVFTGRVFPDAEQSGIQLKEGHVHGKYFELKY 479
C IP+ ++N+++D SI EIFIN+G+KV T RVFP Q+GIQ+ EG G Y+E++Y
Sbjct: 422 CQIPQGHVTLNVYVDNSILEIFINQGQKVLTSRVFPTHGQTGIQVVEGQAFGHYYEMRY 480

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 1161

A DNA sequence (GBSx1237) was identified in *Sagalactiae* <SEQ ID 3603> which encodes the amino acid sequence <SEQ ID 3604>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

45
----- Final Results -----
bacterial cytoplasm --- Certainty=0.2204(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1299-

Example 1162

A DNA sequence (GBSx1238) was identified in *S.galactiae* <SEQ ID 3605> which encodes the amino acid sequence <SEQ ID 3606>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -7.64    Transmembrane  259 - 275 ( 250 - 283)
    INTEGRAL    Likelihood = -4.41    Transmembrane  113 - 129 ( 109 - 130)
    INTEGRAL    Likelihood = -3.03    Transmembrane  180 - 196 ( 180 - 196)
    INTEGRAL    Likelihood = -3.03    Transmembrane  439 - 455 ( 438 - 456)
10  INTEGRAL    Likelihood = -2.81    Transmembrane  298 - 314 ( 298 - 317)
    INTEGRAL    Likelihood = -2.02    Transmembrane  396 - 412 ( 395 - 412)

----- Final Results -----
15  bacterial membrane --- Certainty=0.4057(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

20  >GP:AAC99320 GB:AF059741 sucrose-specific PTS permease [Clostridium
    beijerinckii]
    Identities = 235/453 (51%), Positives = 312/453 (67%), Gaps = 15/453 (3%)

Query: 7   IAKQVINAIGGASNVRSVAHCATRLRVMVKDETVIDKNTVENIEKVQGAFFNSGQYQIIF 66
+AK+++ IGG N++SV HCATRLR+++ D+ I++ +ENI+ V+G FF++ QYQII
25  Sbjct: 6   VAKEILENIGGKENIKSVEHCATRLRLILNDKEKINEKAENIDGVKGQFFSAAQYQIIL 65

Query: 67  GTGTVNKIYDEVVAQGLPTSSTSDQKAEAAKQGNAFQRAIRTFGDVVFVPLPAIVATGLF 126
GTG VN++YD +V Q T + K EA Q Q+ RTFGDVVFV++P +VATGLF
30  Sbjct: 66  GTGCVNEVDVIVGQNSDLV-TGNKKEEAYSQMTLIQKISRFTFGDVVFVPIIPVLVATGLF 124

Query: 127  MGIRGAINNDTVLALFGTTSKAFSSSNFYTYTVVLTDTAFAPFALISWSAFRVFGGNPV 186
MG+RG + N V + NF +T VLTDTAFAP PAL++WS + FGG PV
35  Sbjct: 125  MGLRGLLTNLGVQM-----NENFVLFTQVLTDTAFAPLPAVASTMKKFGGTPV 174

Query: 187  IGLVLGLMMVNSALPNAAVAVASGDAHPIKF--FGF-IPVVGYNQSVLPFAFFVGLLGAKLE 243
IG+V+GLM+V+ +LPNA+AVA+G A PI G IPVVGYNQ SVLPA +G++ AK +
40  Sbjct: 175  IGIIVIGLMLVSPSLPNAYAVAAGTATPINLTILGLNIPVVGYNQSVLPALVLGIIAAKTQ 234

Query: 244  KWLHKKIPDVLDDLVPFLTFTVMSILALFVIGPIPHSVENYVLAGTKFVLNLPLGLSGL 303
K L K +PDVLDL++ PF+T +L L ++GPI H+ E + K + LP GL GL
45  Sbjct: 235  KALKKVVPDVLDLIVTPFITLLFSMVLGLLIVGPIMHNAEQIFGAIKGFMGLPFGGLGGL 294

Query: 304  ILGGVHQIIVVTGVHHIFNLLEAQLIAADGKDPFNAIITAAMTAQAGATLAVGVKTKNKK 363
++GGVHQ+IVVTGVHH N LE +L+++ GKD FNA+IT + AQ A LAV VKTK+KK
50  Sbjct: 295  VVGGVHQLIVVTGVHHALELLELLSSTGKDAFNAMITCGIVAQGAALAVAVKTKDKK 354

Query: 364  LKALAFPAALSAGLGITEPAIFGVNLRFGKPFIMGLIAGAAGGWLASILKLAGTGFGITI 423
++L +A+ A LGITEPAIFGVNLRFPFI G GA GG L+ IL LAGTG GIT
55  Sbjct: 355  KRSLYISSAIPAFLGITEPAIFGVNLRFIKPFIFGCAGGAVGMLSGILHLAGTGMGITA 414

Query: 424  IPGTLILYLNQIVKYLIMVIGTTSLAFVLTVMF 456
+PG LLY+N + Y+++ + ++AF LT F
60  Sbjct: 415  LPGMLLYVN-NLGSYILVNVVAIAVAFCLTLFF 446

```

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3607> which encodes the amino acid sequence <SEQ ID 3608>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
60  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -4.99    Transmembrane  111 - 127 ( 108 - 129)
    INTEGRAL    Likelihood = -4.57    Transmembrane  176 - 192 ( 176 - 193)
    INTEGRAL    Likelihood = -4.35    Transmembrane  436 - 452 ( 431 - 453)
    INTEGRAL    Likelihood = -3.88    Transmembrane  295 - 311 ( 293 - 314)
    INTEGRAL    Likelihood = -3.50    Transmembrane  259 - 275 ( 253 - 277)

```

-1300-

INTEGRAL Likelihood = -2.07 Transmembrane 405 - 421 (405 - 421)
 INTEGRAL Likelihood = -0.43 Transmembrane 219 - 235 (219 - 235)

----- Final Results -----

5 bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAC99320 GB:AF059741 sucrose-specific PTS permease [Clostridium
 beijerinckii]
 Identities = 234/451 (51%), Positives = 312/451 (68%), Gaps = 11/451 (2%)

15 Query: 1 MDNRQIAAEVIEALGGRENVRSAHCASTRLRVMVYDEGKIDKEKAEIDKVKGAFFNSGQ 60
 M + +A E++E +GG+EN++SV HCATRLR+++ D+ KI+++ E ID VKG FF++ Q
 Sbjct: 1 MKEQIVAKEILENIGGKENIKSVEHCASTRRLRLILNDKEKINEKAIENIDGVKQGFSSAAQ 60

20 Query: 61 YQMIFGTGTNNIYDEVVALGLPTSSTSEQKAEAGKHGNIQRAIRTFGDVFVPIIPAI 120
 YQ+I GTG VN +YD +V T K EA + Q+ RTFGDVFVPIIP +V
 Sbjct: 61 YQIILGTGFVNEVYDVIVGQNSDLV-TGNKKEBAYSQMTLIQKISRTFGDVFVPIIPVLV 119

25 Query: 121 ATGLFMGVRGLVTQPAIMDLFGVHEYGENFLMYTRILTDTAFFVLPALVAWSAFRVFGN 180
 ATGLFMG+RGL+T + + ENF+++T++LTDTAFF +LPALVAWS + FGG
 Sbjct: 120 ATGLFMGLRGLLTNLGV-----QMNFVFLFTQVLTDTAFAFLPALVAWSTMKKFGGT 172

30 Query: 181 PIIGIVLGLMLVSNELPNAWVVASGGDVK-PLTFFGF-VPVVGYYQGTVLPFAFFVGLVGAK 238
 P+IGIV+GLMLVS LPNA+ VA+G LT G +PVVGYQG+VLPA +G++ AK
 Sbjct: 173 PVIGIVIGLMLVSPSLPNAYAVAAGTATPINLTILGLNIPVVGYQGSVLPALVLGIIAAK 232

35 Query: 239 LEKWLHKKVPEALDILLVTPFLTFAIMSTLGLFVIGPVFHSLENLVLAGTQAVLHLPFGIA 298
 +K L K VP+ LDL+VTPF+T LGL ++GP+ H+ E L+ + + LPFG+
 Sbjct: 233 TQKALKKKVDPDVLDLIVTPFITLLFSMVLGLLIVGPIMHNAEQIFGAIKGFMGLPFGILG 292

40 Query: 299 GLIVGGIQQILIVVTGIHHIFNFLEAQLIANTGKDPFNAYLTAATAAQAGATLAVAVKTKS 358
 GL+VGG+ QLIVVTG+HH N LE +L+++TGKD FNA +T AQ A LAVAVKTK
 Sbjct: 293 GLVVGGVHQLIVVTGVHHLNALLEVELLSSTGKDAFNAMITCGIVAQGAALAVAVKTKD 352

45 Query: 359 TKLKGAFSTLSALLGITEPAIFGVNLRYPKVFVSLIGGALGGWVAGLFGIAGTGFGI 418
 K + L S + A LGITEPAIFGVNLR+ K F+ G GGA+GG ++G+ +AGTG GI
 Sbjct: 353 KKKRSLYISSAIPAFGLITEPAIFGVNLRFIKPFIFGCAGGAVGMLSGILHLAGTGMGI 412

Query: 419 TVLPGTLLYINGQLLQYLVTMLVGLGVAF 449
 T LPG LLY+N L Y++ +V + VAF +
 Sbjct: 413 TALPGMLLYVN-NLGSYILVNVAIAVAFCL 442

An alignment of the GAS and GBS proteins is shown below.

Identities = 409/618 (66%), Positives = 491/618 (79%), Gaps = 12/618 (1%)

50 Query: 4 NTEIAKQVINAIGGASNVRSVAHCASTRLRVMVKDETVIDKNITVENIEKVQGAFFNSGQYQ 63
 N +IA +VI A+GG NVRSVAHCASTRLRVMV DE IDK E I+KV+GAFFNSGQYQ
 Sbjct: 3 NRQIAAEVIEALGGRENVRSAHCASTRLRVMVYDEGKIDKEKAEIDKVKGAFFNSGQYQ 62

55 Query: 64 IIFGTGTVNKIYDEVVAQGLPTSSTSDQKAEAAKQGNAFQRAIRTFGDVFVPLLPVAT 123
 +IFGTGTVN IYDEVVA GLPTSST+QKAE K GN FQRAIRTFGDVFVP++PAIVAT
 Sbjct: 63 MIFGTGTVNNIYDEVVALGLPTSSTSEQKAEAGKHGNIQRAIRTFGDVFVPIIPVAT 122

60 Query: 124 GLFMGIRGAINNDTVLALFGTTSKAFSSSNFYTYTVVLTDTAFAFFPALISWSAFRVFGG 183
 GLFMG+RG + ++ LFG NF YT +LTDTAFF + PAL++WSAFRVFGG
 Sbjct: 123 GLFMGVRGLVTQPAIMDLFGVHEYG---ENFLMYTRILTDTAFFVLPALVAWSAFRVFGG 179

65 Query: 184 NPVIGLVLGLMMVNSALPNAWVASG-DAHPIKFFGFIPVVGYQNSVLPFAFFVGLGAKL 242
 NP+IG+VLGLM+V++ LPNAW VASG D P+ FFGF+PVVGYQ +VLPFAFFVGL+GAKL
 Sbjct: 180 NPIIGIVLGLMLVSNELPNAWVVASGGDVKPLTFFGFVPVVGYQGTVLPFAFFVGLVGAKL 239

Query: 243 EKWLHKKIPDVLDDLVPFLTFTVMSILALFVIGPIFHSVENYVLAGTKFVLNLPGLSLG 302
 EKWLHKK+P+ LDLL+ PFLT+ +MS L LFVIGP+FHS+EN VLAGT+ VL+LP G++G

-1301-

Sbjct: 240 EKWLHKKVPEALDLLVTPFLTF AIMSTLGLFVIGPVFHSLENLVLAGTQAVLHLFPFGIAG 299

Query: 303 LILGGVHQIIVVTGVHHIFNLLEAQLIAADGKDPFNAIITAAMTAQAGATLAVGVKTKNK 362
LI+GG+ Q+IVVTG+HHIFN LEAQLIA GKDPFNA +TAA AQAGATLAV VKTK+

5 Sbjct: 300 LIVGIGQLIVVTGIIHHIFNLEAQLIANTGKDPFNAYLTAATAAQAGATLAVAVKTKST 359

Query: 363 KLKALAFFAALSAGLGITEPAIFGVNLRFGKPFIMGLIAGAAGGWLASILKLAGTGFGIT 422
KLK LAFF+ LSA LGITEPAIFGVNLR+ K F+ GLI GA GGW+A + +AGTGFGIT

10 Sbjct: 360 KLKGLAFFSTLSALLGITEPAIFGVNLRYPKVFVSGLIGGALGCVAGLFGIAGTGFGIT 419

Query: 423 IIPGTLILYLNQIIVKYLMVIGTTSIAFVLTYMFGYEDKDEKAVAEVSPLVEETDDDDPTI 482
++PGTLLYLNGQ+++YL+ ++ +AF + Y +GY+D++ + V V++T D P +

Sbjct: 420 VLPGTLLYLNGQLLQYLVMTLVGLGVAFAIAYTWGYQDRETLELPAVE--VDQTADQPAL 477

15 Query: 483 TQTSQLRAETIVSPLDGVIALDVTSDPVFSSGIMGDGLAIKPRGNTIYSPVDGQVQIAF 542
+ ET+ SPL+G V+ L VSDPVFSSG MG GLAIKP NT+YSPVDG V+I F

Sbjct: 478 AE-----ETLYSPLNGITVVDLSAVSDPVFSSGAMGQGLAIKPEDNTLYSPVDGKVEIVF 531

20 Query: 543 ETGHAYGIKSDKGAELIHLIGIDTVMNGTCFTSKVKADQKVKKGIDILGTFDSAKIAEAG 602
ETGHAY I S +GAE+L+HIGIDT +M G GF S V Q VKKGD+LG FD +KIAEAG

Sbjct: 532 ETGHAYAITSSQGAEVLLHIGIDTESMAGDGFESLVAVGQAVKKGDLLGHFDPISKIAEAG 591

Query: 603 LDNTAMIIIVTNTADFADV 620
LD+T M+IV+N AD+ V

25 Sbjct: 592 LDDTTMMIVSNIADYQSV 609

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1163

30 A DNA sequence (GBSx1239) was identified in *S.agalactiae* <SEQ ID 3609> which encodes the amino acid sequence <SEQ ID 3610>. This protein is predicted to be fructokinase. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2436(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA02467 GB:D13175 fructokinase [Streptococcus mutans]
Identities = 232/291 (79%), Positives = 257/291 (87%)

45 Query: 1 MTKLYGSIEAGGTFKVCVAVGDEELKVVERMQFPPTTPQETIKKTVDFFKRFEKKLEAVAI 60
M+KLYGSIEAGGTFKVCVAVGDE +++EK+QFPPTTP ETI+KTV FFK+FE L +VAI

Sbjct: 1 MSKLYGSIEAGGTFKVCVAVGDENFQILEKVQFPPTTPYETIEKTVAFFKKFEADLASVAI 60

50 Query: 61 GSFGPIDIDKKSPTYGYITTTPKLHWANVDLLGLISKDFNVPFYFTTVDNSSAYGEVIAR 120
GSFGPIDID+ S TYGYIT+TPK +WANVD +GLISKDF +PFYFTTVDNSSAYGE IAR

Sbjct: 61 GSFGPIDIDQNSDTYGYITSTPKPNWANVDVFLISKDFKIPFYFTTVDNSSAYGETIAR 120

Query: 121 NNIDSLVYYTIGTGIGAGAIQKGEFIGGTGHTTEAGHTYMAHFDQDQANDFKGICPFHNSC 180
+N+ SLVYYTIGTGIGAGAIQ GEFIGG GHTTEAGH YMA HP D + F G CPFH C

55 Sbjct: 121 SNVKSLVYYTIGTGIGAGAIQNGEFIGGMGHTTEAGHVYMAPHPNDVHHGFGVGTCPFHKGK 180

Query: 181 LEGLASGPTLEARTGIRGELIEENSMVWDVQAYYIAQAAIQATVLYRPQVIVFGGGVMAQ 240
LEGLA+GP+LEARTGIRGELIE+NS VWD+QAYYIAQAAIQATVLYRPQVIVFGGGVMAQ

Sbjct: 181 LEGLAAGPSLEARTGIRGELIEQNSEVWDIQAYYIAQAAIQATVLYRPQVIVFGGGVMAQ 240

60 Query: 241 EHMLRRVRQTFATLLNGYLPVPDLSDYIVTPAIEENGSAITLGNFALAKKIS 291

-1302-

EHML RVR+ F +LLN YLPVPD+ DYIVTPA+ ENGSATLGN ALAKKI+
 Sbjct: 241 EHMLNRVREKFTSLNDYLPVPDVKDYIVTPAVAENGSATLGNLALAKKIA 291

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3611> which encodes the amino acid
 5 sequence <SEQ ID 3612>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2012(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 212/293 (72%), Positives = 246/293 (83%)
 Query: 1 MTKLYGSIEAGGTFKFCVAVGDEELKVVEKMQPPTTTTQETIKKTVDDFFKRFEKKLEAVAI 60
 M KLYGSIEAGGTFKFCVAVGDEE VV+K QFPITTP+ETI +T+ +FK FE L +AI
 20 Sbjct: 1 MGKLYGSIEAGGTFKFCVAVGDEEFTVVDKTQFPITTTPEETIARTIAYFKAFEADLAGMAI 60
 Query: 61 GSFGPIDIDKKSKTYGYITTTPKLHWANVDLLGLISKDFNVFFYFTTDVNSSAYGEVIAR 120
 GSFGPIDID S+TYGYITTTPK WANVDLLG +S F +PF TTDVNSSAYGEV+AR
 25 Sbjct: 61 GSFGPIDIDPSSSETYGYITTTPKSGWANVDLLGQLSAAFKIPFDVTTTDVNSSAYGEVLAR 120
 Query: 121 NNIDSLVYYTIGTGIGAGAIQKGEFIGGTGHTGHTYAMMHPQDQANDFKGICPFHNSC 180
 ++SLVYYTIGTGIGAGAIQ G FIGG GHTGHTY+ HP D A F G+CPFH C
 30 Sbjct: 121 PGVESLVYYTIGTGIGAGAIQHGHFIGGLGHTGHTYVMPHPDDMAKGFLGVCPPHKGK 180
 Query: 181 LEGLASGPTLEARTGIRGELIEENSMVWDVQAYYIAQAAIQATVLYRPQVIVFGGGVMAQ 240
 LEG+A+GP++EARTG+RGE +++ + VWD+QA+YIAQAA+QAT+LYRPQVIVFGGGVMAQ
 35 Sbjct: 181 LEGMAAGPSIEARTGVRGERLDQEADVWDIQAFYIAQAALQATMLYRPQVIVFGGGVMAQ 240
 Query: 241 EHMLRRVRQTFTLLNGYLVPDLSYIVTPAIEENGSATLGNFALAKKISK 293
 EHM+ RV F LL+GYLPVPDL+DYIVTPA+ +NGSATLGNFALAK ++G
 Sbjct: 241 EHMVLRVHDKFTALLSGYLPVDLTDYIVTPAVADNGSATLGNFALAKLAAQ 293

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1164

40 A DNA sequence (GBSx1240) was identified in *Sagalactiae* <SEQ ID 3613> which encodes the amino acid sequence <SEQ ID 3614>. This protein is predicted to be Mannosephosphate Isomerase (pmi). Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4717(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA04021 GB:D16594 Mannosephosphate Isomerase [Streptococcus mutans]
 Identities = 232/312 (74%), Positives = 262/312 (83%)

55 Query: 1 MSEPLFLEASMHDKIWGGTKLRDEFGYDIPSETTGEYWAISAHPNGVSRVKNRFGKGCFL 60
 M PLFL++ MH KIWGG +LR EFGYDIPSETTGEYWAISAHPNGVS VKNG +KG L
 Sbjct: 1 MEGPLFLQSQMHKKIWGGNRLRKEFGYDIPSETTGEYWAISAHPNGVSVVKNGVYKGVPL 60

-1303-

Query: 61 DKLYQGEKSLFGNPDPTVFPLLT KILDANDWLSVQVHPDDAYALKHEGELGKTECWYIIS 120
 D+LY + LFGN +VFPLLT KILDANDWLSVQVHPD+AYAL+HEGELGKTECWY+IS
 Sbjct: 61 DELYAEHRELFGNSKSSVFPLLT KILDANDWLSVQVHPDNAYALEHEGELGKTECWYVIS 120

5 Query: 121 ADEGSEIIYGHNAKTKEELRQMIESGDWEHLLTRIPVKSGDFYVPSGTMHAIGKILIL 180
 ADEG+EIIYGH AK+KEELRQMI +GDW+HLLT+IPVK+GDF+YVPSGTMHAIG+GI+IL
 Sbjct: 121 ADEGAEIIYGHEAKSKEELRQMIAAGDWDHLLTKIPVKAGDFFYVPSGTMHAIGRGIMIL 180

10 Query: 181 ETQQSSDITYRVYDFDRPDASGKLRDLHIEQSIDVLTIGKPANTVPANMKLHLSSTLLV 240
 ETQQSSDITYRVYDFDR D G+ R LHIEQSIDVLTIGKPAN PA + L+ L +T+LV
 Sbjct: 181 ETQQSSDITYRVYDFDRKDDQGRKRALHIEQSIDVLTIGKPANATPAWLSLQGLETTVLV 240

Query: 241 SNDFFTIVYKWEISGVTNFKQFAPYLLVSVLDGAGHITVDNKVYTLKKGDHFI LPNDVVKW 300
 S+ FFTIVYKW+ISG +Q APYLLVSVL G G ITV + Y L+KGDH ILPN + W
 15 Sbjct: 241 SSPFFTIVYKWQISGSVKMQQTAPYLLVSVLAGQGRITVGLBQYALRKGDHILIPNTIKSW 300

Query: 301 DIDGQLEIIASH 312
 DG LEIIASH
 20 Sbjct: 301 QFDGDLEIIASH 312

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3615> which encodes the amino acid sequence <SEQ ID 3616>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3714(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

An alignment of the GAS and GBS proteins is shown below.

Identities = 232/312 (74%), Positives = 264/312 (84%)

35 Query: 1 MSEPFLFLEASHMDKIWGGTKLRDEFGYDIPSETTG EYWAI SAHPNGVSRVKNGRFKGCFL 60
 MSEPFLFL+++MHD+IWGGTKLRD F Y+IPS+TTGEYWAI SAHPNGVS V NGR++G L
 Sbjct: 1 MSEPFLFKSTMHDRIWGGTKLRDVFAYNIPSDTTGEYWAI SAHPNGVSTVTNGRYQGQPL 60

40 Query: 61 DKLYQGEKSLFGNPDPTVFPLLT KILDANDWLSVQVHPDDAYALKHEGELGKTECWYIIS 120
 + LY E +LFGN + VFPLLT KILDANDWLSVQVHPDDAY +HEGELGKTECWYIIS
 Sbjct: 61 NTLYAQEPALFGNPKKEVFPPLLT KILDANDWLSVQVHPDDAYGREHEGELGKTECWYIIS 120

45 Query: 121 ADEGSEIIYGHNAKTKEELRQMIESGDWEHLLTRIPVKSGDFYVPSGTMHAIGKILIL 180
 A+EGSEI+YGH AK+KE+LR MIE+G W+ LLTR+PVK+GDF+YVPSGTMHAIGKILIL
 Sbjct: 121 AEEGSEIVYGHQA KSKEDLRAMIEAGAWDDLTRVPVKAGDFFYVPSGTMHAIGKILIL 180

50 Query: 181 ETQQSSDITYRVYDFDRPDASGKLRDLHIEQSIDVLTIGKPANTVPANMKLHLSSTLLV 240
 ETQQSSDITYRVYDFDR D +G LRD LHE+SIDVLTIGKP N+VPA M L ++ +T LV
 Sbjct: 181 ETQQSSDITYRVYDFDRKDVNGNLRDLHIEKSIDVLTIGKPENSVPATMVLDMVATTLV 240

55 Query: 241 SNDFFTIVYKWEISGVTNFKQFAPYLLVSVLDGAGHITVDNKVYTLKKGDHFI LPNDVVKW 300
 S FFTIVYKW S + + KQ APYLLVSVL G G + VD K Y L+KG HFILPNDV W
 Sbjct: 241 STPFFTIVYKWVTSQMVDMKQAAPYLLVSVLKGQGLYVDQKAYELEKGMHFILPNDVKSW 300

Query: 301 DIDGQLEIIASH 312
 DGQLE+I SH
 Sbjct: 301 SFDGQLEMIVSH 312

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1304-

Example 1165

A DNA sequence (GBSx1241) was identified in *S.agalactiae* <SEQ ID 3617> which encodes the amino acid sequence <SEQ ID 3618>. This protein is predicted to be preprotein translocase secA subunit (secA). Analysis of this protein sequence reveals the following:

5 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1102(Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10107> which encodes amino acid sequence <SEQ ID 10108> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA50286 GB:L32090 secA [Listeria monocytogenes]
 Identities = 503/843 (59%), Positives = 643/843 (75%), Gaps = 16/843 (1%)

20 Query: 11 MANILRTVIENDKGELKKLDKIAKKVDSYADHMAALSDEALQAKTPEFKERYQNGETLDQ 70
 MA +L+ + E+ K ++K L++ A ++ + AD AALSD+AL+ KT EFKER Q GETLD
 Sbjct: 1 MAGLLKKIFESGKKDKVYLERKADEIIALADETAALSDDALREKTVEFKERVQKGETLDD 60

25 Query: 71 LLPEAFVAVREASKRVLGLYPYHVQIMGGIVLHHGDIPEMRTGEGKTLTATMPVYLNAIS 130
 LL EAFAV RE +KR LGLYP+ VQ+MGGIVLH +I EM+TGEKTLTAT+PVYLNA+S
 Sbjct: 61 LLVEAFVAVAREGAKRALGLYPFKVQLMGGIVLHEDNIAEMKTGEGKTLTATLPVYLNALS 120

30 Query: 131 GLGVHVITVNEYLSTRDATEMGEVYSWLGLSVGINLAAKSPFEKREAYNCDITYSTNAEV 190
 G GVHV+TVNEYL+ RDA EMG +Y++LGLSVG+NL A S EKREAY CDITYSTN E+
 Sbjct: 121 GEGVHVVTVNEYLAHRDAEEMGVLYNFLGLSVGLNLNALSSTEKREAYACDITYSTNNEL 180

35 Query: 191 GFDYLRDNMVRQEDMVQRPLNYALVDEVDSVLIDEARTPLIVSGPVSSEMNLQYTRADM 250
 GFDYLRDNMNV +E+MVQRPL +A++DEVDS+L+DEARTPLI+SG + + LY RA+
 Sbjct: 181 GFDYLRDNMNVYKEEMVQRPLAFVIDEVDLSILVDEARTPLIISGE-AEKSTILYVRANT 239

40 Query: 251 FVKTL-NSDDYIIDVPTKTIGLSDTGIDKAENYFHLNLYDLENVALTHYIDNALRANYI 309
 FV+TL +DY +D+ TK++ L++ G+ K ENYF + NL+DLEN + H+I AL+ANY
 Sbjct: 240 FVRTLTEEDYTVDIKTKSVQLTEDEGMTKGENYFDVENLFDLENTVILHHIAQALKANYT 299

45 Query: 310 MLLNIDYVVEEQEILIVDQFTGRTMGRRFSDGLHQAIEAKESVPIQEESKTSASITYQ 369
 M L++DYVV ++ E+LIVDQFTGR M+GRRFS+GLHQA+EAKE V IQ ESKT A+IT+Q
 Sbjct: 300 MSLDLDYVV-QDEVLIVDQFTGRIMKGRRFSEGLHQAIEAKEGVTIQNESKTMATITFQ 358

50 Query: 370 NMFRMYHKLAMGTGTGKTEEEEFREIYNMRVPIPTNRPVQRIDHSDLLYPTLDSKFRAY 429
 N FRMY KLAGMTGT KTEEEEFRIYNMRVI IPTN+ + R D DL+Y T+++KF AV
 Sbjct: 359 NYFRMYKLAGMTGTAKTEEEEFRIYNMRVIBIPTNKVIIRDRLDLYTTMEAKFNAV 418

55 Query: 430 VADVKEYEQGQPVLVGTAVETSDLSRKLVAAGVPHEVLNAKNHFKEAQIIMNAGQRG 489
 V D+ ER+ +GQPVLVGTVA+ +LIS KL G+ H+VLNAK H +EA II +AG+RG
 Sbjct: 419 VEDIAERHAKGQPVLVGTAMNI-ELISSKLKRKGIKHDVLNAKQHEREADIKHAGERG 477

60 Query: 490 AVTIATNMAGRGTDIKLGEGVRELGGLCVIGTERHESRRIDNQLRGRSGRQGDPGESQFY 549
 AV IATNMAGRGTDIKLGEG E GGL VIGTERHESRRIDNQLRGRSGRQGDPG +QFY
 Sbjct: 478 AVVIATNMAGRGTDIKLGEGTIEAGGLAVIGTERHESRRIDNQLRGRSGRQGDPGVTQFY 537

65 Query: 550 LSLEDDLMMRRFGTDRIKVVLERMNLAEDDTVIKSKMLTRQVESAQRRVEGNNYDTRKQVL 609
 LS+ED+LMRRFG+D +K ++ER +AED I+SKM++R VESAQRRVEGNN+D+RKQVL
 Sbjct: 538 LSMEDELMMRRFGSDNMKSMMERFGMAED--AIQSKMVSRAVESAQRRVEGNNFDSRKQVL 595

70 Query: 610 QYDDVMREQREIIYANRREVITAERDLGPGLKGMIKRTIKRAVDAHSRSDKNTAA---EA 666
 QYDDV+R+QRE+IY R EVI AE L ++ MI+RT+ V +++ S + A +
 Sbjct: 596 QYDDVLRQQREVIYKQRYEVINAENSLREIIEQMIQRTVNFIVSSNASSHEPEEAWNLAG 655

-1305-

Query: 667 IVNFARSALLDEEAITVSELRLKEAEIKELLYERALAVYEQQIAKLKDPEAIIEFQKVL 726
 I+++ + LL E IT+ +L+ +I+ L+ ++ A Y+++ L PE EF+KV+
 Sbjct: 656 IIDYVDANLLPEGTITLEDLQNRTSEDIQNLILDKIKAAAYDEK-ETLLPPEEFNEFEKVV 714

5 Query: 727 ILMVVDNQWTEHIDALDQLRNSVGLRGYAQNNPIVEYQSEGFRMFQDMIGSIEFDVTRTL 786
 +L VVD +W +HIDA+D LR+ + LR Y Q +P+ EYQSEGF MF+ M+ SI+ DV R +
 Sbjct: 715 LLRVVDTKWVDHIDAMDHLRDGIHLRAYGQIDPLREYQSEGFEMFAMVSSIDEDVARYI 774

10 Query: 787 MKAQIHEQ-ERER-ASQHATTTAEQNISAQHVPMNNEspeyQGIKRNDKPCGSGMKFKN 844
 MKA+I + ERE+ A A AE A+ P+ + Q I RND CPCGSG K+KN
 Sbjct: 775 MKAELRQNLEREQVAKGEAINPAEGKPEAKRQPIRKD----QHIGRNDPCPGSGKKYKN 830

Query: 845 CHG 847
 CHG
 15 Sbjct: 831 CHG 833

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3619> which encodes the amino acid sequence <SEQ ID 3620>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4443(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 710/837 (84%), Positives = 777/837 (92%), Gaps = 3/837 (0%)

30 Query: 11 MANILRTVIENDKGELKKLDKIAKKVDSYADHMAALSDEALQAKTPEFKERYQNGETLDQ 70
 MANILR VIENDKGEL+KL+KIAKKV+SYAD MA+LSD LQ KT EFKERYQ GETL+Q
 Sbjct: 1 MANILRKVIENDKGELRKLEKIAKKVESYADQMASLSDRDLQKLTLEFKERYQKGETLEQ 60

35 Query: 71 LLPEAFVAVREASKRVLGLYPYHVQIMGGIVLHNGDIPEMRTGEGKTLTATMPVYLNAIS 130
 LLPEAFVAVREA+KRVLGL+PY VQIMGGIVLH+GD+PEMRTGEGKTLTATMPVYLNAI+
 Sbjct: 61 LLPEAFVAVREAAKRVLGLFPYRVQIMGGIVLHNGDVPPEMRTGEGKTLTATMPVYLNAIA 120

Query: 131 GLGVHVITVNEYLSTRDATEMGEVYSWLGLSVGINLAAKSPFEKREAYNCDITYSTNAEV 190
 G GVHVITVNEYLSTRDATEMGEVYSWLGLSVGINLAAKSP EKREAYNCDITYSTN+EV
 40 Sbjct: 121 GEGVHVITVNEYLSTRDATEMGEVYSWLGLSVGINLAAKSPAEEKREAYNCDITYSTNSEV 180

Query: 191 GFDYLDRDNMVRQEDMVQRPLN+ALVDEVDVSLIDEARTPLIVSGPVSSSEMNLQYTRADM 250
 GFDYLDRDNMVRQEDMVQRPLN+ALVDEVDVSLIDEARTPLIVSG VSSE NQLY RADM
 45 Sbjct: 181 GFDYLDRDNMVRQEDMVQRPLN+ALVDEVDVSLIDEARTPLIVSGAVSSETNQLYIRADM 240

Query: 251 FVKTLNSDDYIIDVPTKTIGLSDTGIDKAENYFHLNNLYDLENVALTHYIDNALRANYIM 310
 FVKTL S DY+IDVPTKTIGLSD+GIDKAE+YF+L+NLYD+ENVALTH+IDNALRANYIM
 Sbjct: 241 FVKTLTSVDYVIDVPTKTIGLSDSGIDKAESYFNLSNLYDIENVALTHFIDNALRANYIM 300

50 Query: 311 LLNIDYVVSEEQEILIVDQFTGRTMEGRFRSDGLHQAIEAKESVPIQESKTSASITYQN 370
 LL+IDYVVSE+ EILIVDQFTGRTMEGRFRSDGLHQAIEAKE V IQESKTSASITYQN
 Sbjct: 301 LLDIDYVVSEDEILIVDQFTGRTMEGRFRSDGLHQAIEAKGVRIQESKTSASITYQN 360

55 Query: 371 MFRMYHKLAMGTGTGKTEEEEFREIYNMRVPIPIPTNRPVQRIDHSDLLYPTLDSKFRAVV 430
 MFRMY KLAGMTGT KTEEEEFRE+YNMR+IPIPTNRP+ RIDH+DLLYPTL+SKFRAVV
 Sbjct: 361 MFRMYKLAGMTGTAKTEEEEFREVYNMRIPIPTNRPPIARIDHTDLLYPTLESKFRAVV 420

Query: 431 ADVKERYEQGPVLVGTVAVETSDLISRKLVAAGVPHEVLNAKNHFKEAQIIMNAGQRG 490
 DVK R+ +GQP+LVGTVAVETSDLISRKLVAAGVPHEVLNAKNHFKEAQIIMNAGQRG
 60 Sbjct: 421 EDVKTRHAKGPVLVGTVAVETSDLISRKLVEAGIPHEVLNAKNHFKEAQIIMNAGQRG 480

Query: 491 VTIATNMAGRGTDIKLGEGVRELGGLCVIGTERHESRRIDNQLRGRSGRQGDPGESQFYL 550
 VTIATNMAGRGTDIKLGEGVRELGGLCVIGTERHESRRIDNQLRGRSGRQGDPGESQFYL
 65 Sbjct: 481 VTIATNMAGRGTDIKLGEGVRELGGLCVIGTERHESRRIDNQLRGRSGRQGDPGESQFYL 540

-1306-

Query: 551 SLEDDLMMRRFGTDRIKVVLERMNLAEDDTVIKSKMLTRQVESAQRRVEGNNYDTRKQVLQ 610
 SLEDDLMMRRFG+DRIK L+RM L E+DTVIKS ML ROVESAQ+RVEGNNYDTRKQVLQ
 Sbjct: 541 SLEDDLMMRRFGSDRIKAFDRMKLDEEDTVIKSGMLGRQVESAQKRVEGNNYDTRKQVLQ 600

5 Query: 611 YDDVMREQREIIYANRRVITAERDLGPELKGMIKRTIKRAVDAHRSRDKNTAAEAIVNF 670
 YDDVMREQREIIYANRR+VITA RDLGPE+K MIKRTI RAVDAH+RS++ A +AIV F
 Sbjct: 601 YDDVMREQREIIYANRRDVITANRDLGPEIKAMIKRTIDRAVDAHARSNRKDAIDAIVTF 660

10 Query: 671 ARSALLDEEAITVSELRLKEAEIKELLYERALAVYEQQIAKLKDPETAEIEFQKVLILMV 730
 AR++L+ EE I+ ELRGLK+ +IKE LY+RALA+Y+QQ++KL+D EAIEFQKVLILM+
 Sbjct: 661 ARTSLVPEEFISAKELRGLKDDQIKELLYQALAIYDQQLSKLRDQETAEIEFQKVLILMI 720

15 Query: 731 VDNQWTEHIDALDQLRNSVGLRGYAQNNPIVEYQSEGFRMFQDMIGSIEFDVTRTLMKAQ 790
 VDN+WTEHIDALDQLRN+VGLRGYAQNNP+VEYQ+EGF+MFQDMIG+IEFDVTRT+MKAQ
 Sbjct: 721 VDNKWTEHIDALDQLRNAVGLRGYAQNNPVVEYQAEFGKMFQDMIGAIEFDVTRTMMKAQ 780

20 Query: 791 IHEQERERASQHATTAEQNISAQHVPMNNESPEYQGIKRNDKPCGSGMKFKNCHG 847
 IHEQERERASQ ATT A QNI +Q ++ P+ ++RN+ CPCGSG KFKNCHG
 Sbjct: 781 IHEQERERASQRATTAPQNIQSQQSANTDDLPK---VERNEACPCGSGKKFKNCHG 834

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1166

A DNA sequence (GBSx1242) was identified in *S. agalactiae* <SEQ ID 3621> which encodes the amino acid sequence <SEQ ID 3622>. This protein is predicted to be phospho-2-dehydro-3-deoxyheptonate aldolase (aroH). Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3429(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF40753 GB:AE002387 phospho-2-dehydro-3-deoxyheptonate
 aldolase, phe-sensitive [Neisseria meningitidis MC58]
 Identities = 122/348 (35%), Positives = 187/348 (53%), Gaps = 32/348 (9%)

40 Query: 1 MGFHQLSDKINIEILKQKTSLDLEVSQKLAKE-----EELKNIKGEDQRFVLVIV 51
 M H +D I I+ +K+ + + ++KE +E+ +++ G D+R LVI+
 Sbjct: 1 MTHHYPTDDIKIKEVKELLPPIAHLYELPISKEASGLVHRTQREISDLVHGRDKRLVII 60

45 Query: 52 GPCSDADNPKAVLTAKRLAKLEAFAFKDMFLVMRVYTAKPRTNGDGYKGLVHSDKLGVF 111
 GPCS +PKA L YA+RL KL +++++ +VMRVY KPRT G+KGL++ G F
 Sbjct: 61 GPCSIHDPKAALEYAERLLKLRKQYENELLIVMRVYFEKPRIT-VGWKGLINDPHLDGTF 119

50 Query: 112 -----FQARKMHYDIIRETGLLTADELPEMLSVMDLVSYYAIGARSVEDQGHFIS 165
 QAR + + G+ + E L DL+S+ AIGAR+ E Q HR ++
 Sbjct: 120 DINFGRLQARSLLLS-LNNMGMPASTEFLDMITFOYYADLISWGAIGARTTESQVHRELA 178

55 Query: 166 SGIDAPVGMKNPTSGNLRVMFNAVYAAQNQQELFYQNKQ-----VRTDGNLLSHVILRGY 220
 SG+ PVG KN T GNL++ +A+ AA + K V T GN HVILRG
 Sbjct: 179 SGLSCPVGFKNGTGNLKIADAIGAASHSHHFLSVTKAGHSAIVHTGGNPDCHVILRG 238

60 Query: 221 HNADYRSIPNYHYENLLETITHYEETDLQNPFTIVVDTNHDNSGKQFLEQIRIVKSVLADR 280
 PNY E++ E + + +++D +H NS K + Q+ + + + A
 Sbjct: 239 KE-----PNYDAEHVSEAEQLRAAGVTDK-LMIDCSHANSRKDYTRQMEVAQDIAAQL 291

Query: 281 QWHTKIRNYVRGFLIESYLEDRQDKPDVFGKSITDPCLGWDKTEMLI 328
 + + + G ++ES+L +GRQDKP+V+GKSITD C+GW TE L+

-1307-

Sbjct: 292 E---QDGGNIMGVMVESHVLEGRQDKPEVYGKSITDACIGWGATEELL 336

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3623> which encodes the amino acid sequence <SEQ ID 3624>. Analysis of this protein sequence reveals the following:

5 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1171(Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 52/233 (22%), Positives = 93/233 (39%), Gaps = 40/233 (17%)

15 Query: 50 IVGPCSADNPKAVLTYAKRLAKLEAFKDKMFLVMRVYTAKPRTNGDGYKGLVHSDKLG 109
 IVGPCS ++ + A KL + R KPRT+ ++GL
 Sbjct: 19 IVGPCSIESYDHIRLAASSAKKLGYN-----FRGGAYKPRTSAASFQGLG----- 64

20 Query: 110 VFFQARKMHYDIIRETGLLTADELLYPEMLSVMDDLVSYYAIGARSVEDQGHRFISSGID 169
 Q + +++ +E GLL+ E++ L D + +GAR++++ S ID
 Sbjct: 65 --LQGIRYLHEVCQEFGLLSVSEIMSERQLEEAYDYLVDVIQVGARNMQNFEFLKTLSHID 122

25 Query: 170 APVGMKNPTSGNLRVMFNAVYAAQNOQELFYQNKQVRTDGNLLSHVIL--RGYHNADYRS 227
 P+ K + A+ Q+ + S++IL RG D
 Sbjct: 123 KPILFKRGLMATIEEYLGALSYLQDTGK-----SNIILCERGVRGYD--- 164

30 Query: 228 IPNYHYENLLETITHYEETDLQNPFIVVDTNHDNSGKQ-FLEQIRIVKSVLAD 279
 + + +++ ++TDL I+VD +H + L +I K+V A+
 Sbjct: 165 VETRNMLDIMAVPIIQKTDLP---IIVDVSHSTGRDLLLPAAKIAKAVGAN 214

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1167

35 A DNA sequence (GBSx1243) was identified in *S.galactiae* <SEQ ID 3625> which encodes the amino acid sequence <SEQ ID 3626>. This protein is predicted to be AcpS (acpS). Analysis of this protein sequence reveals the following:

40 Possible site: 59
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3620(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG22706 GB:AF276617 acyl carrier protein synthase; AcpS
 [Streptococcus pneumoniae]
 Identities = 61/117 (52%), Positives = 90/117 (76%), Gaps = 1/117 (0%)

50 Query: 1 MIVGHGIDLQEIEAITKAYERNQRFAERVLTEQELLFVKGISNPKRQMSFLTGRWAAKEA 60
 MIVGHGID++E+ +I A R++ FA+RVLIT QE+ F + +RQ+ +L GRW+AKEA
 Sbjct: 1 MIVGHGIDIEELASIESAVTRHEGFQAKRVLTAQEMERFTSLKG-RRQIEYLAGRWSAKEA 59

55 Query: 61 YSKALGTGIGKVNPHDIEILSDDKGAPLITKEPFNGKSFVSISHSGNYAQASVILEE 117
 +SKA+GTGI K+ F D+E+L++++GAP ++ PF+GK ++SISH+ + ASVILEE
 Sbjct: 60 FSKAMGTGISKLGFDLEVLNNERGAPYFSQAPFSGKIWLSISHTDQFVTASVILEE 116

-1308-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3627> which encodes the amino acid sequence <SEQ ID 3628>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2001(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 76/119 (63%), Positives = 99/119 (82%), Gaps = 1/119 (0%)

Query: 1 MIVGHGIDLQEIEAITKAYERNQRFARVLTEQELLFLFKGISNPKRQMSFLTGRWAAKEA 60

MIVGHGIDLQEI AI K Y+RN RFA+++LTEQEL +F+ KR++++L GRW+ KEA

Sbjct: 1 MIVGHGIDLQEISAIEKVYQNRPRFAQKILTEQELAI FESFPY-KRRLNVLGRWSGKEA 59

Query: 61 YSKALGTGIGKVNPHDIEILSDDKGAPLITKEPFNGKSFVSISHSGNYAQASVILEEEK 119

++KA+GTGIG++ F DIEIL+D +G P++TK PF G SF+SISHSGNY QASVILE++K

Sbjct: 60 FAKAIGTGIGRLTFQDIEILNDVRGCPILTKSPFKGNSFISISHSGNYVQASVILEDDK 118

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1168

A DNA sequence (GBSx1244) was identified in *S.agalactiae* <SEQ ID 3629> which encodes the amino acid sequence <SEQ ID 3630>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.24 Transmembrane 78 - 94 (77 - 97)

----- Final Results -----

bacterial membrane --- Certainty=0.2296(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD51027 GB:AF171873 alanine racemase [Streptococcus pneumoniae]

Identities = 227/366 (62%), Positives = 270/366 (73%)

Query: 1 MISSYHRPTRLALIDLEAIANNVKSQVEHIPSDKKTFAVVKANAYGHGAVEVSKYTESIVD 60

M +S HRPT+ALI L AI N++ + HIP AVVKANAYGHGAV V+K I+ VD

Sbjct: 1 MKASPHRPTKALIHLAGAIRQNIQQMGAHIPQGTLLKLVVKANAYGHGAVAVAKAIQDDVD 60

Query: 61 GFCVSNLDEAIELRQAGIVKMILVLGVVMPQVILAKNENITLTVASLEWLRLCQTSAMD 120

GFCVSN+DEAIELRQAG+ K IL+LGV E V LAK + TLTVA LEW++ VD

Sbjct: 61 GFCVSNIDEAIELRQAGLSKPILILGVSEIEAVALAKEYDFTLTVAGLEWIALLDKEVD 120

Query: 121 LSGLEVHIKVDSGMGRIGVRQLDEGNKLISELGESGASVKGIFTHFATADEADNCKFNQ 180

L+GL VH+K+DSGMGRIG R+ E + L + G V+GIFTHFATADE + FN Q

Sbjct: 121 LTGLTVHLKIDSGMGRIGFREASEVEQAQDQLLQGHGVCVEGIFTHFATADEESDDYFNAQ 180

Query: 181 LTFKDFISGLDNCPLVHASNSATSLWHSETIFNAVRLGVVMYGLNPSGTDLDLPYPIN 240

L FK ++ + P+LVHASNSAT+LWH ETIFNAVR+G MYGLNPSG LDLPY +

Sbjct: 181 LERFKTILASMKVPELVHASNSATTLWHVETIFNAVRMGDAMYGLNPSGAVLDLPYDLI 240

Query: 241 PALSLESELVHVHQLHDGSQVGYGATYQVTGDEFVGTVPVIGYADGWTRDMQGFVIVNGE 300

PAL+LES LVHVK + G+ +GYGATYQ ++ + TVPIGYADGWTRDMQ FSV+V+G+

Sbjct: 241 PALTLESALVHVKTVPAGACMGYGATYQADSEQVIATVPVIGYADGWTRDMQNFSVLVDGQ 300

-1309-

Query: 301 LCETIGRVSMQMTIRLPQKYTIGTKVTLIGQQGSCNITTTDVAQKRQTINYEVLCCLSD 360
 C I+GRVSMQ+TIRLP+ Y+GTKVTLIG G IT T VA R TINYEV+CLLSD
 Sbjct: 301 ACPIVGRVSMQITIRLPKLYPLGKVTIGSNGDKETITATQVATYRVNTINYEVVCLLSD 360

5 Query: 361 RIPRY 366
 RIPR Y
 Sbjct: 361 RIPREY 366

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3631> which encodes the amino acid
 10 sequence <SEQ ID 3632>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.34 Transmembrane 82 - 98 (82 - 98)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 The protein has homology with the following sequences in the databases:

>GP:AAD51027 GB:AF171873 alanine racemase [Streptococcus pneumoniae]
 Identities = 222/366 (60%), Positives = 273/366 (73%)

25 Query: 1 MISSFHRPTVARVNLQAIKENVASVQKHIPLGVKTYAVVKADAYGHGAVQVSKALLPQVD 60
 M+S HRPT A ++L AI++N+ + HIP G AVVKA+AYGHGAV V+KA+ VD
 Sbjct: 1 MKASPHRPTKALIHLAGAIRQNIQOMGAHIPQGTLLKLVKANYGHGAVAVAKAIQDDVD 60

30 Query: 61 GYCVSNLDEALQLRQAGIDKEILILGVLLPNELELAVANAITVTIASLDWIALARLEKKE 120
 G+CVSN+DEA++LRQAG+ K ILILGV + LA T+T+A L+WI ++ +
 Sbjct: 61 GFCVSNIDEAIELRQAGLSKPIILGVSEIEAVALAKEYDFTLTVAGLEWIIQALLDKEVD 120

35 Query: 121 CQGLKVHVKVDSGMGRIGLRSSKEVNLLIDSLKELGADVEGIFTHFATADEADDTKFNQQ 180
 GL VH+K+DSGMGRIG R + EV D L++ G VEGIFTHFATADE D FN Q
 Sbjct: 121 LTGLTVHLKIDSGMGRIGFREASEVEQAQDLLQHGVCVEGIFTHFATADEESDDYFNAQ 180

40 Query: 181 LQFFKKLIAGLEDKPRLVHASNSATSIIWSDTIFNAVRLGIVSYGLNPSGSDLSLFPFLQ 240
 L+ FK ++A +++ P LVHASNSAT++WH +TIFNAVR+G YGLNPSG+ L LP+ L
 Sbjct: 181 LERFKTILASMKEVPPELVHASNSATTLWHEVETIFNAVRMGDAMYGLNPSGAVLDLPYDLI 240

40 Query: 241 EALSLESSLVHVKMISAGDTVGYGATYTAKKSEYVGTVPIGYADGWTRNMQGFSVLVDGQ 300
 AL+LES+LVHV K + AG +GYGATY A + + TVPIGYADGWTR+MQ FSVLVDGQ
 Sbjct: 241 PALTLESALVHVKTVPAGACMGYGATYQADSEQVIATVPIGYADGWTRDMQNFSVLVDGQ 300

45 Query: 301 FCEIIGRVSMQMTIRLPKAYPLGKVTIGSNQKNISTTDIANYRNTINYEVLCCLSD 360
 C I+GRVSMQ+TIRLPK YPLGKVTIGSN K I+ T +A YR TINYEV+CLLSD
 Sbjct: 301 ACPIVGRVSMQITIRLPKLYPLGKVTIGSNGDKETITATQVATYRVNTINYEVVCLLSD 360

50 Query: 361 RIPRIY 366
 RIPR Y
 Sbjct: 361 RIPREY 366

An alignment of the GAS and GBS proteins is shown below.

Identities = 247/366 (67%), Positives = 295/366 (80%)

55 Query: 1 MISSYHRPTRALIDLEAIANNVKSQVEHIPSDKKTFAVVKANAYGHGAVEVSKYIESIVD 60
 MISS+HRPT A ++L+AI NV SVQ+HIP KT+AVVKA+AYGHGAV+VSK + VD
 Sbjct: 1 MISSFHRPTVARVNLQAIKENVASVQKHIPLGVKTYAVVKADAYGHGAVQVSKALLPQVD 60

60 Query: 61 GFCVSNLDEAIELRQAGIVKMILVLGVVMPPEQVILAKNENITLTVASLEWLRLCQTSADV 120
 G+CVSNLDEA++LRQAGI K IL+LGV++P ++ LA IT+T+ASL+W+ L + +
 Sbjct: 61 GYCVSNLDEALQLRQAGIDKEILILGVLLPNELELAVANAITVTIASLDWIALARLEKKE 120

Query: 121 LSGLEVHIKVDSGMGRIGVRQLDEGNKLISELGESGASVKGIFTHFATADEADNCKFNQQ 180
 GL+VH+KVDSGMGRIG+R E N LI L E GA V+GIFTHFATADEAD+ KFNQQ

-1310-

Sbjct: 121 CQGLKVHVKVDSGMGRIGLRSSKEVNLLIDSLKELGADVEGIFTFATADEADDTKFNQQ 180

Query: 181 LTFFKDFISGLDNCPLVHASNSATSLWHSETIFNAVRLGVVMYGLNPSGTDLDLPYPIN 240
L FFK I+GL++ P LVHASNSATS+WHS+TIFNAVRLG+V YGLNPSG+DL LP+P+

5 Sbjct: 181 LQFFKKLIAGLEDKPRLVHASNSATSIIWSDTIFNAVRLGIVSYGLNPSGSDLSLPFPLQ 240

Query: 241 PALSLESELVHVKQLHDGSQVGYGATYQVTGDEFVGTVPVIGYADGWTRDMQGFSVIVNGE 300
ALSLES LVHVK + G VGYGATY E+VGTVPVIGYADGWTR+MQGFSV+V+G+

10 Sbjct: 241 EALSLESSLVHVKMISAGDTVGYGATYTAKKSEYVGTVPVIGYADGWTRNMQGFSVLVDGQ 300

Query: 301 LCEIIGRVSMQDMTIRLPQKYTIGTKVTLIGQQGSCNITTTDVAQKRQTINYEVLCCLSD 360
CEIIGRVSMQ+TIRLP+ Y +GTKVTLIG NI+TTD+A R TINYEVLCCLSD

Sbjct: 301 FCEIIGRVSMQDLTIRLPKAYPLGTVTLIGSNQQKNISTTDIANYRNTINYEVLCCLSD 360

15 Query: 361 RIPRYY 366
RIPR Y

Sbjct: 361 RIPRIY 366

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
20 vaccines or diagnostics.

Example 1169

A DNA sequence (GBSx1245) was identified in *S.galactiae* <SEQ ID 3633> which encodes the amino acid sequence <SEQ ID 3634>. This protein is predicted to be immunogenic secreted protein precursor. Analysis of this protein sequence reveals the following:

25 Possible site: 27
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
30 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 1988.

A related GBS gene <SEQ ID 8745> and protein <SEQ ID 8746> were also identified. Analysis of this
35 protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 4
McG: Discrim Score: 8.81
GvH: Signal Score (-7.5): 0.659999
Possible site: 27
40 >>> Seems to have a cleavable N-term signal seq.
ALOM program count: 0 value: 1.06 threshold: 0.0
PERIPHERAL Likelihood = 1.06 247
modified ALOM score: -0.71

45 *** Reasoning Step: 3

----- Final Results -----
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50 bacterial cytoplasm --- Certainty=0.0000(Not Clear)

SEQ ID 8746 (GBS98) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 5; MW 80kDa).

GBS98-His was purified as shown in Figure 192, lane 9.

-1311-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1170

A DNA sequence (GBSx1246) was identified in *S.agalactiae* <SEQ ID 3635> which encodes the amino acid sequence <SEQ ID 3636>. This protein is predicted to be junction specific DNA helicase (mmsA) (recG). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -0.16      Transmembrane      530 - 546 ( 530 - 546)

----- Final Results -----
      bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA90280 GB:Z49988 MmsA [Streptococcus pneumoniae]
Identities = 483/671 (71%), Positives = 568/671 (83%)

Query: 1   MLLQSPISNLKGFPGKSAEKFKQLDIYTVEDLLLYYPFRYEDFKSKSVFDLVDGEKAVIT 60
          M L P+ L G GPKSAEK+ KL I ++DLLLY+PFRYEDFK+K V +L DGEKAV++
Sbjct: 1   MNLHQPLHVLPGVGPKSAEKYAKLGIENLQDLLLYFPFRYEDFKTKQVLELEDGEKAVLS 60

Query: 61  GLVVTTPANVQYVGFKRNRLSFKLRQGEAVLNVSFFNQPYLADKIELGQEVAVFGKWDATK 120
          G VVTPA+VQYVGFKRNRL F L+QGE V V+FFNQPYLADKIELG +AVFGKWD K
Sbjct: 61  GQVVTTPASVQYVGFKRNRLRFSKLQGEVVFVAVNFFNQPYLADKIELGATLAVFGKWDRAK 120

Query: 121 SAITGMKVLAQVEDDMQPVYRVAQGISQSTLIKAISAFEISAHLELKENLPATLLEKYR 180
          +++TGMKVLAQVEDD+QPVYR+AQGISQ++L+K IK+AF+ L ++ENLP +LL+KY+
Sbjct: 121 ASLTGMKVLAQVEDDLQPVYRLAQGISQASLVKVIKTAFTDQGLDLLIEENLPQSLDLKYK 180

Query: 181 LMGRSQACLAMHFPKDIITYKQALRRIKFEELFYFQMNLQVLKSENKSETNGLPILYSKH 240
          LM R QA AMHFPKD+ EYKQALRRIKF ELFYFQM LQ LKSEN+ + +GL + +S+
Sbjct: 181 LMSRCQAVRAMHFPKDLAEYKQALRRIKFAELFYFQMQLQTLKSENVRVQSGSLVNLWSQE 240

Query: 241 AMETKISSLPFILTNQKRSLEILSDMSSGAHMNRLQGDVGSCKTVIAGLSMYAAYTA 300
          + +SLPF LT AQ++SL EIL+DM S HMNRLQGDVGSCKTV+AGL+M+AA TA
Sbjct: 241 KVTAVKASLPFALTQAQEKSLQEILTDMKSDHMHNRLQGDVGSCKTVVAGLAMFAAVTA 300

Query: 301 GFQSALMVPTTEILAEQHYISLQELFPDLSTAILTSGMKAQVKTVLAAIANGSVDIMVGT 360
          G+Q+ALMVPTTEILAEQH+ SLQ LFP+L +A+LT +KAA KR VL IA G D+I+GT
Sbjct: 301 GYQALMVPTTEILAEQHFESLQNLFPNLKIALLTGSLKAAEKREVLETIAKGEADLIIGT 360

Query: 361 HALIQDSVQYHKLGLVITDEQHRFGVKQRRIFREKGENPDVLMMTATPIPTLTAITAFGE 420
          HALIQD V+Y +LGL+I DEQHRFGV QRR I REKG+NPDLMMTATPIPTLTAITAFG+
Sbjct: 361 HALIQDGEYARLGLIIDEQHRFGVGQRRILREKGNPDVLMMTATPIPTLTAITAFGD 420

Query: 421 MDVSIIDELPAGRKPIITRWVKHEQLGTVLEWVKGEIQKDAQVYVISPLIEESEALDLKN 480
          MDVSIID++PAGRKPI+TRW+KHEQL VL W++GE+QK +Q YVISPLIEESEALDLKN
Sbjct: 421 MDVSIIDQMPAGRKPIVTRWIKHEQLPQVLTWLEGEIQKGSQAYVISPLIEESEALDLKN 480

Query: 481 AVALHAELSTYFEGIAKVALVHGRMKNDEKDAIMQDFKDKKSHILVSTTVIEVGNNVPNA 540
          A+AL EL+T+F G A+VAL+HGRMK+DEKD IMQDFK++K+ ILVSTTVIEVGNNVPNA
Sbjct: 481 AIALSEELTTHFAGKAEVALVHGRMKSEKQIMQDFKERTDILVSTTVIEVGNNVPNA 540

Query: 541 TIMIIMDADRFGLSQLHQLRGRVGRGYKQSYAVLVANPKTDSGKKRMTIMTETTDGFVLA 600
          T+MIIMDADRFGLSQLHQLRGRVGRG KQSYAVLVANPKTDSGK RM IMTETT+GFVLA
Sbjct: 541 TVMIIMDADRFGLSQLHQLRGRVGRGDKQSYAVLVANPKTDSGKDRMRIMTETTINGFVLA 600

Query: 601 ESDLKMRGSGEIFGTRQSGIPEFQVADIVEDYPILIEARRVASDIVKDNWNKENTEWALI 660
          E DLKMRGSGEIFGTRQSG+PEFQVADI+ED+PILEEAR+VAS I W+E+ EW +I
Sbjct: 601 EEDLKMRGSGEIFGTRQSGLEPFQVADIIEDFPILIEARKVASYISSIEAWQEDPEWRMI 660
```

-1312-

Query: 661 LDNLRQHSDFD 671

+L + D

Sbjct: 661 ALHLEKKEHLD 671

5

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3637> which encodes the amino acid sequence <SEQ ID 3638>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

10

INTEGRAL Likelihood = -0.16 Transmembrane 530 - 546 (530 - 546)

----- Final Results -----

15

bacterial membrane --- Certainty=0.1065(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 641/671 (95%), Positives = 655/671 (97%)

20

Query: 1 MLLQSPISNLKGFGPKSAEKFKQLDIYTVEDLLLYYPFRYEDFKSKSVFDLVDGEKAVIT 60

M+L +P+SNLKGFGPKSAEKFKQLDIYTVEDLLLYYPFRYEDFKSKSVFDLVDGEKAVIT

Sbjct: 1 MILTAPMSNLKGFGPKSAEKFKQLDIYTVEDLLLYYPFRYEDFKSKSVFDLVDGEKAVIT 60

25

Query: 61 GLVVTTPANVQYYGFKRNRLSFKLRQGEAVLNVSFFNQPYLADKIELGQEVAVFGKWDATK 120

GLVVTTPANVQYYGFKRNRLSFKLRQGEAVLNVSFFNQPYLADKIELGQEVAVFGKWDATK

Sbjct: 61 GLVVTTPANVQYYGFKRNRLSFKLRQGEAVLNVSFFNQPYLADKIELGQEVAVFGKWDATK 120

30

Query: 121 SAITGMKVLAQVEDDMQPVYRVAQGSIQSSTLIKAISAFEISAHLELKENLPATLLEKYR 180

SAITGMKVLAQVEDDMQPVYRVAQGSIQSSTLIKAISAFEI AHLELKENLPATLLEKYR

Sbjct: 121 SAITGMKVLAQVEDDMQPVYRVAQGSIQSSTLIKAISAFEIDAHELELKENLPATLLEKYR 180

35

Query: 181 LMGRSQACLAMHFPKDITEYKQALRRIKFEELFYFQMNQVLKSENKSETNGLPILYSKH 240

LMGRSQACLAMHFPKDITEYKQALRRIKFEELFYFQMNQVLK+ENKSETNGLPILYSK

Sbjct: 181 LMGRSQACLAMHFPKDITEYKQALRRIKFEELFYFQMNQVLKAENKSETNGLPILYSKR 240

40

Query: 241 AMETKISSLPFILTNQAKRSLDEILSDMSSGAHMNRLLQGDVGSGKTVIAGLSMYAAYTA 300

AMETKISSLPFILTNQAKRSLD+ILSDMSSGAHMNRLLQGDVGSGKTVIAGLSMYAAYTA

Sbjct: 241 AMETKISSLPFILTNQAKRSLDDILSDMSSGAHMNRLLQGDVGSGKTVIAGLSMYAAYTA 300

45

Query: 301 GFQSALMVPTTEILAEQHYISLQELFPDLSIAILTSGMKAARKTVLAAIANGSVDIMVGT 360

GFQSALMVPTTEILAEQHYISLQELFPDLSIAILTSGMKAARKTVLAAIANGSVDIMVGT

Sbjct: 301 GFQSALMVPTTEILAEQHYISLQELFPDLSIAILTSGMKAARKTVLAAIANGSVDIMVGT 360

50

Query: 361 HALIQDSVQYHKLGLVITDEQHRFGVKQRRIFREKGENPDVLMMTATPIPTLTAITAFGE 420

HALIQDSVQYHKLGLVITDEQHRFGVKQRRIFREKGENPDVLMMTATPIPTLTAITAFGE

Sbjct: 361 HALIQDSVQYHKLGLVITDEQHRFGVKQRRIFREKGENPDVLMMTATPIPTLTAITAFGE 420

55

Query: 421 MDVSIIDELPAGRKPIITRWVKHEQLGTVLEWVKGELQKDAQVYVISPLIEESEALDLKN 480

MDVSIIDELPAGRKPI+TRWVKHEQLGTVLEWVKGELQKDAQVYVISPLIEESEALDLKN

Sbjct: 421 MDVSIIDELPAGRKPIITRWVKHEQLGTVLEWVKGELQKDAQVYVISPLIEESEALDLKN 480

60

Query: 481 AVALHAELSTYFEGIAKVALVHGRMKNDEKDAIMQDFKDKKSHILVSTTVIEGVNVPNA 540

AVALHAELSTYFEGIAKVALVHGRMKNDEKDAIMQDFKDKKSHILVSTTVIEGVNVPNA

Sbjct: 481 AVALHAELSTYFEGIAKVALVHGRMKNDEKDAIMQDFKDKKSHILVSTTVIEGVNVPNA 540

65

Query: 541 TIMIIMDADRFGLSQLHQLRGRVGRGYKQSYAVLVANPKTDSGKKRMTIMTETTDGFFVLA 600

TIMIIMDADRFGLSQLHQLRGRVGRGYKQSYAVLVANPKTDSGKKRMTIMTETTDGFFVLA

Sbjct: 541 TIMIIMDADRFGLSQLHQLRGRVGRGYKQSYAVLVANPKTDSGKKRMTIMTETTDGFFVLA 600

Query: 601 ESDLKMRGSGEIFGTRQSGIPEFQVADIVEDYPILIEEARRVASDIVKDNWKNTEWALI 660

ESDLKMRGSGEIFGTRQSGIPEFQVADIVEDYPILIEEAR+V++ IV D NW +W L+

Sbjct: 601 ESDLKMRGSGEIFGTRQSGIPEFQVADIVEDYPILIEEARKVSAIVSDPNWIYEKQWQLV 660

Query: 661 LDNLRQHSDFD 671

N+R+ +D

-1313-

Sbjct: 661 AQNIRKKEVVD 671

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1171

A DNA sequence (GBSx1247) was identified in *S.agalactiae* <SEQ ID 3639> which encodes the amino acid sequence <SEQ ID 3640>. This protein is predicted to be aryl-alcohol dehydrogenase (b1647). Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1562(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10105> which encodes amino acid sequence <SEQ ID 10106> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

20 >GP:BAB07646 GB:AP001520 aryl-alcohol dehydrogenase [Bacillus halodurans]
    Identities = 173/300 (57%), Positives = 224/300 (74%)

Query: 7 IGQTGIQATRIALGCMRMSDLKKGQAEEVVGTTALDLGINFFDHADIYGGGLSELRFDAI 66
      +G + ++ +A+GCMR++ + K+AE V TAL+ G NFFDHADIYGGG E F DAI
25 Sbjct: 6 LGSSSLEVPVAVGCMRINAIKKEAERFVQTALEQGANFFDHADIYGGGECEEIFADAI 65

Query: 67 KHLNVNRDKMIIQSKCGIREGYFDFSKEYILSSVDGILERLGTEYLDLFLILHRPDVLVEP 126
      + R+K+I+QSKCGIREG FDFSKEYIL SVDGIL+RL T+YLD L+LHRPD LVEP
30 Sbjct: 66 QMNEAVREKIIILQSKCGIREGRFDFSKEYILQSVGDILQRLKTDYLDVLLLHRPDALVEP 125

Query: 127 EEVAAEFTKLRAEGKVKHFGVSNQNRFMELLQSYLDEPLAVNQLQLSPAHTPMFDAGLN 186
      EEVAAEF L + GKV+HFGVSNQN Q+ELL+ ++ +P+ NQLQLS + M +G+N
35 Sbjct: 126 EEVAAFDLLESSGKVRHFGVSNQNPQMIELLKKFVRQPIVANQLQLSITNATMISSGIN 185

Query: 187 VNMLNKASIEHDDGIVDYICRLKRVTIQAWSPFQIDLSRGLFVNHPDYKELNETIAKLAKN 246
      VNM N+++I D ++DYCRL VTIQ WSPFQ G+F+ + + ELN+ I +LA+
40 Sbjct: 186 VNMENESAINRDGSLDYICRLHDVTIQPWSPPFYGFEGVFLGNDLFPELNKKIDELA EK 245

Query: 247 YNVSSEAIIVIAWILRHPAKMQAIVGSMNPSRLKAIDKANDIALTRKEWYDIYRSAGNILP 306
      Y VS+ I IAW+LRHPA MQ ++G+MN RLK KA++I LTR+EWY+IYR+AGNILP
40 Sbjct: 246 YEVSNTTIAIAWLLRHPANMQPVIGTMNLKRLKDCCKASEIRLTREEWYIYRAAGNILP 305

```

There is also homology to SEQ ID 780.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1172

A DNA sequence (GBSx1248) was identified in *S.agalactiae* <SEQ ID 3641> which encodes the amino acid sequence <SEQ ID 3642>. This protein is predicted to be shikimate 5-dehydrogenase (aroE) (aroE). Analysis of this protein sequence reveals the following:

```

50 Possible site: 21
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

-1314-

bacterial cytoplasm --- Certainty=0.0988(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC74762 GB:AE000264 putative oxidoreductase [Escherichia coli K12]
 Identities = 114/279 (40%), Positives = 171/279 (60%), Gaps = 3/279 (1%)

10 Query: 10 LTGLIANPARHSLSPMLMNTSFQEKMNMYAYLTFEVEEGKLTAVRGVRLGIRGVNVSM 69
 L GL+A P RHSLSP M N + ++ + + Y+ FEV+ A+ G++AL +RG VSM
 Sbjet: 9 LIGLMAYPIRHSLSPEMQNKALEKAGLPFTYMAFEVDNDSFPGAIEGLKALKMRGTGVSM 68

15 Query: 70 PFKQSVIPLDDLSPOAKLVGAVNTIVNQCGTGRVLGHMTDGIGCFKALAAQGFSAKNKI 129
 P KQ +D+L+P AKLVGA+NTIVN G R G+ TDG G +A+ GF K K
 Sbjet: 69 PNKQLACEYVDELTPAAKLVGAINTIVNDDGYLR--GYNTDGTGHIRAIKESGFDIKGKT 126

20 Query: 130 ITIAGIGGSGKAVAVQAAMEGVAEIRLFNRNSSNYDKVIDLSDKIKKQFQIKVVVDYLEN 189
 + + G GG+ A+ Q A+EG+ EI+LFNR +DK + + ++ + V V L +
 Sbjet: 127 MVLLGAGGASTAIGAQAIEGLKEIKLFNRDEFFDKALAFARVNENTDCVVTVTDLAD 186

25 Query: 190 KTAFKDAIRTSHFYIDATSLGMRPLDNYSLINDPEILTPNLVVVDLVYKPKETALLRFVR 249
 + AF +A+ ++ + T +GM+PL+N SL+ND +L P L+V + VY P T LL+ +
 Sbjet: 187 QQAFABALASADILTNGTKVGMKPLENESLVNDISLLHPGLLVTECVYNPHMTKLLQQAQ 246

Query: 250 QNGVKHAYNGLGMLIYQGAEAFQLITNQEMPISSVERVL 288
 Q G K +G GML++QCAE F L T ++ P+ V++V+
 Sbjet: 247 QAGCK-TIDGYGMLLWQGAEQFTLWTGKDFPLEYVKQVM 284

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3643> which encodes the amino acid
 sequence <SEQ ID 3644>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have an uncleavable N-term signal seq

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

40 >GP:AAC74762 GB:AE000264 putative oxidoreductase [Escherichia coli]
 Identities = 132/280 (47%), Positives = 186/280 (66%), Gaps = 3/280 (1%)

45 Query: 11 LVSLLATPIRHSLSPKMHNEAYAKGLDYAYLAFEVGTQLADAVQGIRALGIRGSNVSM 70
 L+ L+A PIRHSLSP+M N+A K GL + Y+AFEV + A++G++AL +RG+ VSM
 Sbjet: 9 LIGLMAYPIRHSLSPEMQNKALEKAGLPFTYMAFEVDNDSFPGAIEGLKALKMRGTGVSM 68

50 Query: 71 PNKEAILPLDDLSPAEELVGAVNTVNVKDGKGLHVGHTDGIGALRALADEGVSVKNKI 130
 PNK+ +D+L+PAA+LVGA+NT+VN DG +L G+ TDG G +RA+ + G +K K
 Sbjet: 69 PNKQLACEYVDELTPAAKLVGAINTIVNDDG--YLRGYNTDGTGHIRAIKESGFDIKGKT 126

55 Query: 131 ITLAGVGGAGKAIQVLAFAFDGAKEVRLFNQATRLSSVOKLVTKLNQLTRTKVTLQDLED 190
 + L G GGA AI Q A +G KE++LFNR+ ++N+ T VT+ DL D
 Sbjet: 127 MVLLGAGGASTAIGAQAIEGLKEIKLFNRDEFFDKALAFARVNENTDCVVTVTDLAD 186

60 Query: 191 QTAFKEAIRESHLFIDATSVGMKPLENLSLITDPELIRPDLVVFIDIVYSPAETKLLAFAR 250
 Q AF EA+ + + + T VGMKPLEN SL+ D L+ P L+V + VY+P TKLL A+
 Sbjet: 187 QQAFABALASADILTNGTKVGMKPLENESLVNDISLLHPGLLVTECVYNPHMTKLLQQAQ 246

Query: 251 QHGAQKVINGLGMVLYQGAEAFKLITGQDMPVDAIKPLLG 290
 Q G K I+G GM+L+QCAE F L TG+D P++ +K ++G
 Sbjet: 247 QAGC-KTIDGYGMLLWQGAEQFTLWTGKDFPLEYVKQVMG 285

An alignment of the GAS and GBS proteins is shown below.

-1315-

Identities = 166/288 (57%), Positives = 221/288 (76%)

Query: 4 LNGETLLTGLIANPARHSLSPIMWNITSFQEKNNMYAYLTFEVEEGKLTEAVRGVRLGIR 63
 L+G TLL L+A P RHSLSP M N ++ + ++YAYL FEV +L +AV+G+RALGIR
 5 Sbjet: 5 LSGHTLLVSLLATPIRHSLSPKMHNEAYAKLGLDYAYLAFEVGTQLADAVQGIRALGIR 64

Query: 64 GNVNSMPFKQSVIPLDDLSLQAKLVGAVNTIVNQGGRVLGHMTDGIGCFKALAAQGF 123
 G NVSMP K++++PLDDLSL A+LVGAVNT+VN+ G G LVGH+TDGIG +ALA +G
 10 Sbjet: 65 GSNVSMFNKEAILPLDDLSLPAELVAVNTIVNKGKGLVGHITDGIGALRALADEGV 124

Query: 124 SAKNKIITITAGIGSGKAVAVQAAMEGVAEIRLFNRNSSNYDKVIDLSDKIKKQFQIKVV 183
 S KNKIIT+AG+GG+GKA+AVQ A +G E+RLFNK ++ V L K+ + + KV
 15 Sbjet: 125 SVKNKIITLAGVGGAGKAIQVQLAFDGAKEVRLFNQATRLSSVQKLVTKLNLTRTKVT 184

Query: 184 VDYLENKTAFKDAIRTSHFYIDATSLGMRPLDNYSLINDPEILTPNLVVVDLVYKPKETA 243
 + LE++TAFK+AIR SH +IDATS+GM+PL+N SLI DPE++ P+IVV D+VY P ET
 20 Sbjet: 185 LQDLEDQTAFKEAIRESHLFIDATSVGMKPLENLSLITDPELIRPDLVVDIVYSPAETK 244

Query: 244 LLRFVRQNGVKHAYNGLGMLIYQGAEAFQLITNQEMPISSVERVLQTE 291
 LL F RQ+G + NGLGM++YQGAEAF+LIT Q+MP+ +++ +L E
 25 Sbjet: 245 LLAFARQHGAKVINGLGMVLYQGAEAFKLITGQDMPVDAIKPFLGDE 292

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1173

A DNA sequence (GBSx1249) was identified in *S.agalactiae* <SEQ ID 3645> which encodes the amino acid sequence <SEQ ID 3646>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 30 INTEGRAL Likelihood = -6.16 Transmembrane 57 - 73 (53 - 76)

----- Final Results -----
 bacterial membrane --- Certainty=0.3463(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1174

A DNA sequence (GBSx1250) was identified in *S.agalactiae* <SEQ ID 3647> which encodes the amino acid sequence <SEQ ID 3648>. Analysis of this protein sequence reveals the following:

Possible site: 17
 45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2333(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10103> which encodes amino acid sequence <SEQ ID 10104> was also identified.

-1316-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05343 GB:AP001512 L-asparaginase [Bacillus halodurans]

Identities = 158/319 (49%), Positives = 214/319 (66%), Gaps = 4/319 (1%)

5 Query: 1 MKKILVLHTGGTISMNANEKGQVMSSADNPMKYVDLSLDDL-DLTVVDFLNLPSPQITPH 59
MKK+LV+HTGGTI+M+ +EKG V NP+ SL + + V DFLN+PSP +TP
Sbjct: 1 MKKVLVIHTGGTIAMHEDEKGGVQPKETNPLFATVESLTSIASIEVDDFLNIPSPHMTPE 60

10 Query: 60 HMLDIYHYLKQHASN--FDGVVITHGTDLEETAYFLDTMILPKIPIIITGAMRSTNELG 117
M + LK N FDGVVITHGTDLEETAY LD ++ ++P+++TGAMRS+NELG
Sbjct: 61 LMFQLAERLKS RVGNESFDGVVITHGTDLEETAYLLDLLLDWEVPVVVTGAMRSSNELG 120

15 Query: 118 SDGVVNYLSALRVANSTKAADKGVLVVMNDEIHAAKYVTKTHTTNVSTFQTPHGLGII 177
+DG +N++SA++ A + +A KGVLVV NDEIH AK VTKTHT+NV+TFQ+P +GP+GI+
Sbjct: 121 ADGPHNFISAVKTAATDEAKGKGVLVVFNDEIHTAKNVTKTHTSNVATFQSPQYGPIGIV 180

20 Query: 178 MKQDLLFFKATEERVRFDLKITGTVPVIVKAYAGMGDSGIISFLNSQNISGLVIEALGAG 237
K+ + F A + + + I V ++KAYAGM D ++ + I GLVIEA G G
Sbjct: 181 TKRGVTFHHAPS YKESYTVSSIDHRVLLKAYAGM-DGSVVDIAIDTGIDGLVIEAFGQG 239

25 Query: 238 NMPPKAAQEIEELIEQGVFVVLVSRFCNGIAEPVYEGGGAKLQESGMFVKELNAPKA 297
N+PP I+ L + +PVVLVSR +GI + Y YEGGG L++ GV+F LN KA
Sbjct: 240 NLPPAVVPSIKRLHQANIPVVLVSRVSGIVQETIYAYEGGGRHLKDLGVIFTNGLNGQKA 299

25 Query: 298 RLKLLIALNAGLTGQNLKD 316
RLKLL+AL + L++
Sbjct: 300 RLKLLVALELTDRKKLQE 318

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3649> which encodes the amino acid sequence <SEQ ID 3650>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.28 Transmembrane 245 - 261 (243 - 261)

35 ----- Final Results -----

bacterial membrane --- Certainty=0.1914(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the databases:

>GP:BAB05343 GB:AP001512 L-asparaginase [Bacillus halodurans]

Identities = 158/320 (49%), Positives = 218/320 (67%), Gaps = 5/320 (1%)

45 Query: 1 MKKILVLHTGGTISMQADNSGRVVPNODNPM-TKIHAQAQDIQLTVSDFLNLPSPHITPH 59
MKK+LV+HTGGTI+M D G V P + NP+ + + + V DFLN+PSPH+TP
Sbjct: 1 MKKVLVIHTGGTIAMHEDEKGGVQPKETNPLFATVESLTSIASIEVDDFLNIPSPHMTPE 60

50 Query: 60 HMLSIVHHIQERT--DVFDGIVITHGTDLEETAYFLDTMALPTNIPVVLITGAMRSSNEV 117
M + ++ R + FDG+VITHGTDLEETAY LD + L +PVV+TGAMRSSNE+
Sbjct: 61 LMFQLAERLKS RVGNESFDGVVITHGTDLEETAYLLDLL-LDWEVPVVVTGAMRSSNEL 119

55 Query: 118 GSDGIYNYLTALRVASSDKAKEKGVLVVMNDEIHAAKYVTKTHTTNISTFQTPHGLGI 177
G+DG +N+++A++ A++D+AK KGVLVV NDEIH AK VTKTHT+N++TFQ+P +GP+GI
Sbjct: 120 GADGPHNFISAVKTAATDEAKGKGVLVVFNDEIHTAKNVTKTHTSNVATFQSPQYGPIGI 179

60 Query: 178 IMKNDLLFFKTAEPRIRFDLRCSGTIPIIKAYAGMGDSILSLTPGSIQGLVIEALGA 237
+ K + F + + + I + ++KAYAGM DGS++ + I GLVIEA G
Sbjct: 180 VTKRGVTFHHAPS YKESYTVSSIDHRVLLKAYAGM-DGSVVDIAIDTGIDGLVIRAFGQ 238

60 Query: 238 GNVPPPLAVGEIEHLIALGIPVILVSRFCNGMAEPVYAYEGGGAMLQAGVMFVKELNAPK 297
GN+PP V I+ L IPV+LVSR +G+ + YAYEGGG L++ GV+F LN K
Sbjct: 239 GNLPPAVVPSIKRLHQANIPVVLVSRVSGIVQETIYAYEGGGRHLKDLGVIFTNGLNGQK 298

Query: 298 ARLKLLIALNAGLTGQELKD 317

-1317-

ARLKL+AL ++L++
 Sbjct: 299 ARLKLVALELTDRKKLQE 318

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 242/321 (75%), Positives = 275/321 (85%), Gaps = 1/321 (0%)

Query: 1 MKKILVLHTGGTISMNANEKGQVMSSADNPMKYVDLSLDDDLTVVDFLNLPSQITPHH 60
 MKKILVLHTGGTISM A+ G+V+ + DNPM + + D+ LTV DFLNLPS ITPHH
 10 Sbjct: 1 MKKILVLHTGGTISMQADNSGRVVPNQDNPMTKIHAAQDIQLTVSDFLNLPSPHITPHH 60

Query: 61 MLDIYHYLKQHASNFDGCVITHGTDTLEETAYFLDTMILP-KIPIITGAMRSTNELGSD 119
 ML IYH++++ FDG+VITHGTDTLEETAYFLDTM LP IP+++TGAMRS+NE+GSD
 15 Sbjct: 61 MLSIYHHIQERTDVFVGIVITHGTDTLEETAYFLDTMALPTNIPVVLTGAMRSSNEVGSD 120

Query: 120 GVYNYLSALRVANSTKAADKGVLVVMNDEIHAACYVTKTHTTNVSTFQTPTHGPLGIIMK 179
 G+YNYL+ALRVA+S KA +KGVLVVMNDEIHAACYVTKTHTTN+STFQTPTHGPLGIIMK
 20 Sbjct: 121 GIYNYLTALRVASSDKAKEKGVLVVMNDEIHAACYVTKTHTTNISTFQTPTHGPLGIIMK 180

Query: 180 QDLLFFKATEERVRFDLDKITGTPIVKAYAGMGDSGIISFLNSQISGLVIEALGAGNM 239
 DLLFFK E R+RFDL I+GT+PI+KAYAGMGD I+S L +I GLVIEALGAGN+
 25 Sbjct: 181 NDLLFFKTAEPRIRFDLRCSGTPIIKAYAGMGDSILSLTPGSIQGLVIEALGAGNV 240

Query: 240 PPAAQIEELIEQGVFVVLVSRFCNGIAEPVYEGGGAQLQESGVMFVKELNAPKARL 299
 PP A BIE LI G+PV+LVSRFCNG+AEPVY YEGGGA LQE+GVMFVKELNAPKARL
 30 Sbjct: 241 PPLAVGIEHLIALGIPVILVSRFCNGMAEPVYAYEGGGAQLQESGVMFVKELNAPKARL 300

Query: 300 KLLIALNAGLTGQNLKDYIEG 320
 KLLIALNAGLTGQ LKDYIEG
 35 Sbjct: 301 KLLIALNAGLTGQELKDYIEG 321

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1175

35 A DNA sequence (GBSx1251) was identified in *Sagalactiae* <SEQ ID 3651> which encodes the amino acid sequence <SEQ ID 3652>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4427(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:CAB85142 GB:AL162757 conserved hypothetical protein [Neisseria meningitidis Z2491]
 Identities = 87/285 (30%), Positives = 138/285 (47%), Gaps = 35/285 (12%)

50 Query: 4 KAVFFDIDGTLNDRKNVQKSTIK-AIRNLKDQGILVGLATGRG----PSFVQPFLLENLG 58
 K VFFDID TL + + ++K A+ L+ +GIL LATGR P V+ + G
 Sbjct: 11 KIVFFDIDDTLYRKVTDTLRFSVKTAVALRGKILATATGRSLATIPKVRDMMATG 70

Query: 59 LDFAVTYNGQYIYSRSEIYTNQLSKTTVYRLIRYAGARRREISLGTASGLLGSGIIGLG 118
 +D VT NQ+ + + + R+ + SLG +G G+
 55 Sbjct: 71 MDVAVTTINGQFALLHGKTVCVEPMDAGLMGRVCAHLD-----SLGMDYAFVVGGE--GIA 122

Query: 119 TSRLGQIVSSLVPRKWAKAIERSFKHFIRRIKPNIDSLMVILREPIYQVVLVATEGE-- 176
 S L + V R+ KH I +P+YQ+++ A E E
 60 Sbjct: 123 VSALSECVC-----RALKH-----IASDFFADKDYFSSKPVYQMLVFAEENEMP 166

Query: 177 --SERIQKQFPRVKL/TRSSPYSMDEVISEGQSKVKGIERVGRYGFDLSEVIAFGDSDNDI 234

-1318-

S+ ++++ +K R ++D++ G SK GI V + G ++++V+AFGD ND+
 Sbjct: 167 LWSDIVERE--GLKTVRWHEEAVDLLPAGASKTDGIRSVVEALGLEMADVMAFGDGLNDV 224

5 Query: 235 EMLSQVGIGVAMGNASQQVRENARYTTADNNDGISKALAHYGLI 279
 EMLS+VG GVAMGN Q +E A+Y ++DG+ + L G+I
 Sbjct: 225 EMLSEVGFVAMGNGEQAQAAKAYVCPGVDEGVLRLQLDGLVI 269

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3653> which encodes the amino acid sequence <SEQ ID 3654>. Analysis of this protein sequence reveals the following:

10 Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 15 bacterial cytoplasm --- Certainty=0.6014(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 320/459 (69%), Positives = 391/459 (84%)

20 Query: 1 MAIKAVFFDIDGTLNDRKNVQKSTIKAIRNLKQGGILVGLATGRGSPFVQPFLENLGLD 60
 + +KAVFFDIDGTLNDRKN+QK+T KAI+ LK QGI+VGLATGRGP FVQPFLEN GLD
 Sbjct: 1 LTVKAVFFDIDGTLNDRKNIQKTTQKAIQQLKKQGIMVGLATGRGPGFVQPFLENFGLD 60

25 Query: 61 FAVTYNGQYIYSRSEIIYTNQLSKTTVYRLIRYAGARRREISLGTASGLLGSGIIGLGS 120
 FAVTYNGQYI +R +++Y NQL K+ +Y++IRYA ++REISLGTASGL GS II +GTS
 Sbjct: 61 FAVTYNGQYILTRDKVLYQNQLPKSMIYKVIYANEKKREISLGTASGLAGSRIIDMGTS 120

30 Query: 121 RLQGVSSSLVPRKWAKAIERSFKHFIRRIKPQNIDSLMVILREPIYQVVLVATEGESERI 180
 GQ++SS VP+ WA+ +E SFKH IRRIKPQ+ +L+ I+REPIYQVVLVA++ E+++I
 Sbjct: 121 PFGQVISSFPKSWARTVEGSPFKHLIRRIKPQSFERNLVTIMREPIYQVVLVASQAETKKI 180

35 Query: 181 QKQFPRVKLTRSSPYSDVISEGQSKVKGIERVQGRYGFDLSEVIAFGDSDNDIEMLSQV 240
 Q++FP +K+TRSSPYSD+D+IS QSK+KGIER+G+ +GFDLSEV+AFGDSDND+EMLS V
 Sbjct: 181 QEKFPHIKITRSSPYSLDLISVDQSKIKGIERLGEMFGFDLSEVMAFGDSDNDLEMLSGV 240

40 Query: 241 GIGVAMGNASQQVRENARYTTADNNDGISKALAHYGLIQFEIEKTFSSRDENFNKVKSF 300
 GIG+AMGNA V++ A +TT NN+DGISKALAHYGLI F+IEK+F SRDENFNKVK F
 Sbjct: 241 GIGIAMGNAETVVKDGAHFTTDSNNNDGISKALAHYGLIHFDIEKSFKSRDENFNKVKDF 300

45 Query: 301 HLLMDGETIETPRLYDSKEAGFRSDFKVEEIVEFLYAASQGNQKVFDDQSIRNLHLAIDKA 360
 H LMD +TIETPR Y EAG+RS FKVEEIVEFLYAAS+G+Q+ F Q+I +LH A+D+A
 Sbjct: 301 HRLMDSDTIETPRSYTISEAGYRSGFKVEEIVEFLYAASKGDQQFTQAI FDLHGAVDQA 360

50 Query: 421 HFDPVTHKILKPDWEEHFAPEPSIRRELDSSQIQKSLNR 459
 HFDPVTHKI KPD W+E APE +I++ELD Q+QKSL R
 Sbjct: 421 HFDPVTHKIQKPDYWQERHAEVAIKKELDKQLQKSLQR 459

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 1176

A DNA sequence (GBSx1252) was identified in *S.agalactiae* <SEQ ID 3655> which encodes the amino acid sequence <SEQ ID 3656>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence

60

-1319-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1671(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

A related GBS nucleic acid sequence <SEQ ID 10101> which encodes amino acid sequence <SEQ ID 10102> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:BA06903 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 61/141 (43%), Positives = 92/141 (64%)

Query: 22 YERILVAIDGSTESLAFEKAVNVALRNDSELILTHVIDTRALQSFATFDITYIEKLEKE 81
 Y ILVA+DGST+++ A KA N A ++L + HVID+R+ + +D + E +
 15 Sbjct: 2 YNHILVAVDGSTQAKRALYKAFNYAKEFKADLFICHVIDSRSFATVEQYDRTVVGAAELD 61

Query: 82 AKDVL EEYEKQAREKGADKVRQVIEFGNPKTLLAHDIPKEKVDLIMVGATGLNTERFX 141
 K +L+ Y ++A + G DKV +++FG+PK ++ I +K +DLI+ GATGLN ERF
 Sbjct: 62 GKLLQRYSEEAEKAGVDKVHTILDFGSPKANISKTAQKYDIDLITGATGLNAVERFL 121

20 Query: 142 IGSSSEYILRHAKVDLLIVRD 162
 +GS SE + RHAK D+LIVR+
 Sbjct: 122 MGSVSESVARHAKCDVLIVRN 142

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3657> which encodes the amino acid sequence <SEQ ID 3658>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1296(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 117/156 (75%), Positives = 135/156 (86%)

Query: 12 LEEDRLMSQKYERILVAIDGSTESLAFEKAVNVALRNDSELILTHVIDTRALQSFATFD 71
 L+ED MS KY+RILVAIDGS ESELA F K VNVALRND+ L+L HVIDTRALQS ATFD
 40 Sbjct: 25 LKEDSSMSLKYKRILVAIDGSYESELA FNKGVNVALRNDATLLLVHVIDTRALQSVATFD 84

Query: 72 TYIEKLEKEAKDVL EEYEKQAREKGADKVRQVIEFGNPKTLLAHDIPKEKVDLIMVGA 131
 TYIEKLE+EA KDV L+++EKQA+ G ++Q+IEFGNPK LLAHDIP++E DLIMVGA
 Sbjct: 85 TYIEKLEQEAKDVLDDFEKQAQIAGITNIKQIIEFGNPKNLLAHDIPDRENADLIMVGA 144

45 Query: 132 TGLNTERFXIGSSSEYILRHAKVDLLIVRDPNKT 167
 TGLNTER IGSSSEYI+RHAK+DLL+VRD KT+
 Sbjct: 145 TGLNTERLLIGSSSEYIMRHAKIDLLVVRDSTKTL 180

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1177

A DNA sequence (GBSx1253) was identified in *S.galactiae* <SEQ ID 3659> which encodes the amino acid sequence <SEQ ID 3660>. This protein is predicted to be aspartate aminotransferase (aspC). Analysis of this protein sequence reveals the following:

55 Possible site: 47
 >>> Seems to have no N-terminal signal sequence

-1320-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2803(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAC21948 GB:U32714 aminotransferase [Haemophilus influenzae Rd]
 Identities = 142/212 (66%), Positives = 181/212 (84%), Gaps = 1/212 (0%)

Query: 1 MKIFDKSMKLEHVAYDIRGPVLEEADMRANGEKILRLNTGNPAAFGFEAPDEVIRDLIT 60
 M++F KS KLEHV YDIRGPV +EA R+ G KIL+LN GNPA FGFEAPDE++ D++
 Sbjct: 1 MRLFPKSDKLEHVICYDIRGPVHKALRLEEENKILKLNIGNPAPFGFEAPDEILVDVLR 60

15 Query: 61 NARESEGYSDSKGIFSARKAVMQYYQLQNI-HVDMDDIYIVNGVSEGISMSMQALLDND 119
 N ++GY DSKG++SARKA++QYYQ + I ++D+YI NGVSE I+M+MQALL++ D
 Sbjct: 61 NLPASQGYCDSKGLYSARKAIVQYYQSKGILGATVNDVYIGNGVSELITMAMQALLNDGD 120

20 Query: 120 EVLVPMPDYPLWTACVSLAGGNAVHYICDEEANWYPDIDDIKSKITSKTKAIVLINPNNP 179
 EVLVPMPDYPLWTA V+L+GG AVHY+CDE+ANW+P IDDIK+K+ +KTKAIV+INPNNP
 Sbjct: 121 EVLVPMPDYPLWTAAVTLGGKAVHYLCDEBANWFPTIDDIKAKVNAKTKAIVINPNNP 180

25 Query: 180 TGAVYPREILQEIVDIARQNDLIIFSDEYDR 211
 TGAVY +E+LQEIV+IARQN+LIIF+DE+YD+
 Sbjct: 181 TGAVYSKELLQEIVEIARQNNLIIFADEIYDK 212

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3661> which encodes the amino acid sequence <SEQ ID 3662>. Analysis of this protein sequence reveals the following:

30 Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2936(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 170/212 (80%), Positives = 193/212 (90%), Gaps = 1/212 (0%)

40 Query: 1 MKIFDKSMKLEHVAYDIRGPVLEEADMRANGEKILRLNTGNPAAFGFEAPDEVIRDLIT 60
 MKI +KS KLEHVAYDIRGPVL+EA+RM A+GEKILRLNTGNPAAFGFEAPDEVIRDLI
 Sbjct: 13 MKIIEKSSKLEHVAYDIRGPVLDEANRMIASGEKILRLNTGNPAAFGFEAPDEVIRDLIV 72

45 Query: 61 NARESEGYSDSKGIFSARKAVMQYYQLQNI-HVDMDDIYIVNGVSEGISMSMQALLDND 119
 NAR SEGYSDSKGIFSARKA+MQY QL+ VD++DIY+ NGVSE IS+S+QALLDN D
 Sbjct: 73 NARLSEGYSDSKGIFSARKAIMQYCQLKGFPDVEDIYILGNGVSELISISLQALLDNGD 132

50 Query: 120 EVLVPMPDYPLWTACVSLAGGNAVHYICDEEANWYPDIDDIKSKITSKTKAIVLINPNNP 179
 EVLVPMPDYPLWTACVSL GG AVHY+CDEEA WYPDI DIKSKITS+TKAIV+INPNNP
 Sbjct: 133 EVLVPMPDYPLWTACVSLGGGKAVHYLCDEEAGWYPDIADIKSKITSRTKAIVVINPNNP 192

55 Query: 180 TGAVYPREILQEIVDIARQNDLIIFSDEYDR 211
 TGA+YP+EIL++IV +AR++ LIIF+DE+YDR
 Sbjct: 193 TGALYPKEILEDIVALAREHQLIIFADEIYDR 224

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1178

60 A DNA sequence (GBSx1254) was identified in *S.agalactiae* <SEQ ID 3663> which encodes the amino acid sequence <SEQ ID 3664>. Analysis of this protein sequence reveals the following:

-1321-

Possible site: 60

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood =-14.75 Transmembrane 38 - 54 (29 - 60)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.6901(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 10 A related GBS nucleic acid sequence <SEQ ID 9389> which encodes amino acid sequence <SEQ ID 9390> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3665> which encodes the amino acid sequence <SEQ ID 3666>. Analysis of this protein sequence reveals the following:

15 Possible site: 43

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood =-15.97 Transmembrane 35 - 51 (25 - 58)

20 ----- Final Results -----

bacterial membrane --- Certainty=0.7389(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 25 An alignment of the GAS and GBS proteins is shown below.

Identities = 51/87 (58%), Positives = 63/87 (71%), Gaps = 7/87 (8%)

30 Query: 1 MAKKPWEKKVVENNSHRKDKITRTSRGVVSSTPWITAFLSAFFVIVVAILFIVFYTSNRG 60
 MAK+PWE+K+V++ + TR SR STPW+TA LS FFVI+VAILFI FYTSN G
 Sbjct: 1 MAKEPWEKIVDDTIGTR---TRKSRNAFISTPWLITALLSVFFVIIIVAILFIFFYTSNSG 57

Query: 61 EDRAKETSGFYGASSQKVNSSKTKKAS 87
 +R ET+GFYGAS+ K KT+KAS

35 Sbjct: 58 SNRQAETNGFYGASTHK----KTRKAS 80

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1179

- 40 A DNA sequence (GBSx1255) was identified in *S.agalactiae* <SEQ ID 3667> which encodes the amino acid sequence <SEQ ID 3668>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0815(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3669> which encodes the amino acid sequence <SEQ ID 3670>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

-1322-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.0107(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 43/64 (67%), Positives = 53/64 (82%)

10 Query: 1 MKVALIPEKCIACGLCQTYSNIFDYQDDGIVKFSDTDNLEKEIPSSDQDTVLAVKSCPTK 60
 MKV++IPEKCIACGLCQTY++FDY D+GIV FS + + I SD+D +LAVKSCPTK
 Sbjct: 1 MKVSIPEKCIACGLCQTYSSLFDYHDNGIVTFSSSSETSQSICPSDKDAILAVKSCPTK 60

15 Query: 61 ALTI 64
 ALT+
 Sbjct: 61 ALTL 64

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 **Example 1180**

A DNA sequence (GBSx1256) was identified in *S.agalactiae* <SEQ ID 3671> which encodes the amino acid sequence <SEQ ID 3672>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
 25 INTEGRAL Likelihood = -10.61 Transmembrane 47 - 63 (41 - 69)

----- Final Results -----

30 bacterial membrane --- Certainty=0.5246(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC36851 GB:L23802 pore-forming peptide [Enterococcus faecalis]
 Identities = 42/130 (32%), Positives = 63/130 (48%), Gaps = 9/130 (6%)

35 Query: 7 KIRYHWQPELSWAIYWSIAIPIFIGLSLLYERTE---IPSQVFLFAIFIVLVGIGFH 63
 K +++WQPEL+ IIYWS +FI L L E I + V V F +F L G
 Sbjct: 3 KQKFYWQPELASTIIYWSCTFCILFISLILALENNGPYLISNLVMVPFFVFAYL---GIA 59

40 Query: 64 RYFVIEEDGYLRIVSFNLRRTKFPIEDIAKIEVTKSSVTIKFNNNHE--RIFYMRKWPK 121
 R F + E L + + R+ P+ I K+ + S+ I + E ++F M+K
 Sbjct: 60 RSFNMETS-LIVRDVLWFRKKALPLSQIEKVITYNEKSIEIFSSEFKEGSKVFLMKKTD 118

45 Query: 122 KYFLDALAIE 131
 FL+AL I+
 Sbjct: 119 SLFLEALKIK 128

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3673> which encodes the amino acid sequence <SEQ ID 3674>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

50 INTEGRAL Likelihood = -9.87 Transmembrane 47 - 63 (41 - 69)
 INTEGRAL Likelihood = -3.35 Transmembrane 20 - 36 (18 - 37)

55 ----- Final Results -----

 bacterial membrane --- Certainty=0.4949(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1323-

The protein has homology with the following sequences in the databases:

```
>GP:AAC36851 GB:L23802 pore-forming peptide [Enterococcus faecalis]
Identities = 42/130 (32%), Positives = 70/130 (53%), Gaps = 12/130 (9%)

5  Query: 7  KIRYHWQPELSWSIIYWSIAFAPIFVGLSLLYERTE---IPSRVFILFAIFAVLVGIGLH 63
      K +++WQPEL+ +IIYWS F +F+ L L E I + V + F +FA L G+
      Sbjct: 3  KQKPHYWQPELASTIIYWSCTFCILFISLILALENNGPYLISNLVMPFFVFAYL---GIA 59

10  Query: 64 RYF-IIENNGILRIVSFKLFGPRKLLISTITKIEVTKSTLCL--HVEDKSYLFYMRKWP 119
      R F + E + I+R V + F + L +S I K+ + + + ++ S +F M+K
      Sbjct: 60 RSNMTETSLIVRDVLW--FRKKALPLSQIEKVITYNEKSIEIFSSSEFKEGSKVFLMKKKT 117

      Query: 120 KKYFLDALAV 129
      FL+AL +
15  Sbjct: 118 DSLFLEALKI 127
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 115/162 (70%), Positives = 132/162 (80%), Gaps = 1/162 (0%)

20  Query: 1  MIKLF GKIRYHWQPELSWAIYWSIAIAPIFIGLSLLYERTEIPSQVFLFAIFVLVGI 60
      MIKLF GKIRYHWQPELSW+IIYWSIA APIF+GLSLLYERTEIPS+VF+LFAIF VLVGI
      Sbjct: 1  MIKLF GKIRYHWQPELSWSIIYWSIAFAPIFVGLSLLYERTEIPSQVFLFAIFAVLVGI 60

25  Query: 61 GFHRYFVIEEDGYLRIVSFNFLRRTKFPEDIAKIEVTKSSVTIKFNNNHERIFYMRKWP 120
      G HRYF+IE +G LRIVSF K I I KIEVTKS++ + + +FYM RKWP
      Sbjct: 61 GLHRYFIIENNGILRIVSFKLFGPRKLLISTITKIEVTKSTLCLHVEDK-SYLFYMRKWP 119

      Query: 121 KKYFLDALAIEPTFKGEVELLDNLIKMDYFECYRYDKKALTK 162
      KKYFLDALA+ P F+GEV L DN IK+DYFE Y++DKKALT+
30  Sbjct: 120 KKYFLDALAVNPYFQGEVILSDNFIKLDYFEVYQHDKKALTR 161
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1181

35 A DNA sequence (GBSx1257) was identified in *S.galactiae* <SEQ ID 3675> which encodes the amino acid sequence <SEQ ID 3676>. This protein is predicted to be peptidase t (pepT). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

40  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2913(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA20627 GB:L27596 tripeptidase [Lactococcus lactis]
Identities = 274/406 (67%), Positives = 334/406 (81%), Gaps = 4/406 (0%)

50  Query: 1  MSYEKLLERFLTYVKINTRSNPNSTQTPTTQSQVDFALTVLKPEMEAIGLKD VHYLPSNG 60
      M YEKLL RFL YVK+NTRS+ NST TP+TQ+ V+FA + +M+A+GLKDVHYL SNG
      Sbjct: 1  MKYEKLLPRFLEYVKVNTSRDENSTTTPTSTQALVEFAHK-MGEDMKALGLKDVHYLESNG 59

55  Query: 61 YLVGTLPATSDRLRHKIGFISHMDTADFNAENITPQIVDYKGGD--IELGDSGYILSPKD 118
      Y++GT+PA +D+ KIG ++H+DTADFNAE + PQI++ G+ I+LGD+ + L PKD
      Sbjct: 60 YVIGTIPANTDKKVRKIGLLAHLDTADFNAEGVNPQILENYDGESVIQLGDTFTLDPKD 119

      Query: 119 FPNLNNYHGQTLITTDGKTLTGADDKSGIAEIMTAMEYLAS-HPEIEHCEIRVGFGPDDE 177
      FPNL NY GQTL+ TDG TLLG+DDKSG+AEIMT +YL + +P+ EH EIRVGFGPDDE
```

-1324-

Sbjct: 120 FPNLKNYKQTLVHTDGTTLGSDDKSGVAEIMTLADYLLNINPDFEHGEIRVGFPGPDEE 179

Query: 178 IGIGADKFDVKDFDVFAYTVDDGGLGELQYETFSAGLELTFEGRNVHPGTAKNQMINA 237
IG+GADKFDV DFDVDFAYTVDDGGLGELQYETFSAG + F+G+NVHPGTAKN M+NA

5 Sbjct: 180 IGVGADKFDVADFDVDFAYTVDDGGLGELQYETFSAGAVIEFQGNVHPGTAKNMMVNA 239

Query: 238 LQLAMDFHSQLPENERPEQTDGYQGFFHYLYDLSTVDQAKSSYIIRDFEEVDLKRKHLA 297
LQLA+D+H+ LPE +RPE+T+G +GF+HL L GT ++A++ YIIRD EE F +RK L

10 Sbjct: 240 LQLAIDYHNALPEFDRPEKTEGREGFFHLLKLDGTPEEARAQYIIRDEEGKFNERKALM 299

Query: 298 QDIADNMNEALQSERVKVLYDQYNNMKVIEKDMTPINIAKEVMEELDIKPIIEPIRG 357
Q+IAD MN L RVK + DQYNNM ++IEKDM+ I+IAK+ ME LDI PIIEPIRG

Sbjct: 300 QEIADKMNAELGQNRVKPVIKDQYNNMAQIIEKDMSIIDIAKAMENLDIPIIEPIRG 359

15 Query: 358 TDGSKISFMGIPTPNLFAGGENMHGRFEFVSLQIMEKAVDVILGIV 403
TDGSKISFMG+PTPNLFAGGENMHGRFEFVS+QIMEKAVD +L I+

Sbjct: 360 TDGSKISFMGLPTPNLFAGGENMHGRFEFVSQIMEKAVDTLLEII 405

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3677> which encodes the amino acid sequence <SEQ ID 3678>. Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2938 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 305/406 (75%), Positives = 352/406 (86%), Gaps = 1/406 (0%)

Query: 1 MSYEKLLERFLTYVKINTRSNPNSTQIPTTQSQVDFALITVLKPEMEAIGLKDVHYLPSNG 60
M Y+ LL+RF+ YVK+NTRS P+S TP+T+SQ FALT+LKPEMEAIGL+DVHY P NG

35 Sbjct: 5 MKYDNLDRFIKYVKVNTSRVDPSETTPSTESQEAFAITLKPMEEAIGLQDVHYNPVNG 64

Query: 61 YLVGTLPATSDRLRHKIGFISHMDTADFNAENITPQIVD-YKGGDIELGDSGYILSPKDF 119
YL+GTLPA + L KIGFI+HMDTADFNAEN+ PQI+D Y+GGDI LG S Y L PK F

Sbjct: 65 YLIGTLPANNTLTKIGFIAHMDTADFNAENVNPQIIDNYQGGDITLGSSNYKLDPKAF 124

40 Query: 120 PNLNNYHGQTLITTDGKTLGADDKSGIAEIMTAMEYLASHPETEHCEIRVGFPGPDEEIG 179
PNLNNY GQTLITTDG TLLGADDKSGIAEIMTA+E+L S P+IEHC+I+V FGPDEEIG

Sbjct: 125 PNLNNYIGQTLITTDGTLGADDKSGIAEIMTAIEFLTSQPQIEHCIDIKVAFGPDEEIG 184

45 Query: 180 IGADKFDVKDFDVFAYTVDDGGLGELQYETFSAGLELTFEGRNVHPGTAKNQMINALQ 239
+GADKF+V DF+VDFAYT+DGGPLGELQYETFSAA LE+TF GRNVHPGTAK+QMINAL+

Sbjct: 185 VGADKFEVADFEVDFAITMDGGPLGELQYETFSAAALEVTFLGRNVHPGTAKDQMINALE 244

Query: 240 LAMDFHSQLPENERPEQTDGYQGFFHYLYDLSTVDQAKSSYIIRDFEEVDLKRKHLAQD 299
LA+DFH +LP +RPE TDGYQGFFHYL L+GTV++A++SYIIRDFEE F RK ++

50 Sbjct: 245 LAIDFHEKLPKDRPEYTDGYQGFFHYLTGLTGTVVEEARASYIIRDFEEASFARKVKVEN 304

Query: 300 IADNMNEALQSERVKVLYDQYNNMKVIEKDMTPINIAKEVMEELDIKPIIEPIRG 359
IA +MN L ++RV V+L DQYNNMKVIEKDMT I +AKEVMEEL IKP+IEPIRG

Sbjct: 305 IAQSMNAQLGTRVVLVELNDQYNNMKVIEKDMTAEI LAKEVMEELAIKPVIEPIRG 364

55 Query: 360 GSKISFMGIPTPNLFAGGENMHGRFEFVSLQIMEKAVDVILGIVAK 405
GSKISFMGIPTPN+AGGENMHGRFEFVSLQIME+AVDVI+G+V K

Sbjct: 365 GSKISFMGIPTPNIFAGGENMHGRFEFVSLQIMERAVDVIIGLVCK 410

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1325-

Example 1182

A DNA sequence (GBSx1258) was identified in *Sagalactiae* <SEQ ID 3679> which encodes the amino acid sequence <SEQ ID 3680>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 43
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -12.26    Transmembrane  481 - 497 ( 477 - 508)
      INTEGRAL    Likelihood = -9.45     Transmembrane  510 - 526 ( 506 - 534)
      INTEGRAL    Likelihood = -7.96     Transmembrane  316 - 332 ( 310 - 334)
      INTEGRAL    Likelihood = -7.54     Transmembrane  354 - 370 ( 351 - 373)
10   INTEGRAL    Likelihood = -7.11     Transmembrane  385 - 401 ( 383 - 409)
      INTEGRAL    Likelihood = -6.58     Transmembrane  215 - 231 ( 211 - 233)
      INTEGRAL    Likelihood = -6.48     Transmembrane  71 - 87 ( 69 - 91)
      INTEGRAL    Likelihood = -6.32     Transmembrane  110 - 126 ( 106 - 133)
      INTEGRAL    Likelihood = -5.10     Transmembrane  446 - 462 ( 443 - 465)
15   INTEGRAL    Likelihood = -3.29     Transmembrane  418 - 434 ( 418 - 435)
      INTEGRAL    Likelihood = -2.55     Transmembrane  263 - 279 ( 263 - 279)
      INTEGRAL    Likelihood = -2.02     Transmembrane  142 - 158 ( 141 - 159)
      INTEGRAL    Likelihood = -1.70     Transmembrane  184 - 200 ( 184 - 200)

20   ----- Final Results -----
      bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

25 A related GBS nucleic acid sequence <SEQ ID 8747> which encodes amino acid sequence <SEQ ID 8748> was also identified. Analysis of this protein sequence reveals the following:

```

      Lipop: Possible site: -1    Crend: 6
      McG: Discrim Score:      -10.58
      GvH: Signal Score (-7.5): -1.1
30   Possible site: 32
   >>> Seems to have no N-terminal signal sequence
      ALOM program    count: 13 value: -12.26 threshold: 0.0
      INTEGRAL    Likelihood = -12.26    Transmembrane  470 - 486 ( 466 - 497)
      INTEGRAL    Likelihood = -9.45     Transmembrane  499 - 515 ( 495 - 523)
35   INTEGRAL    Likelihood = -7.96     Transmembrane  305 - 321 ( 299 - 323)
      INTEGRAL    Likelihood = -7.54     Transmembrane  343 - 359 ( 340 - 362)
      INTEGRAL    Likelihood = -7.11     Transmembrane  374 - 390 ( 372 - 398)
      INTEGRAL    Likelihood = -6.58     Transmembrane  204 - 220 ( 200 - 222)
      INTEGRAL    Likelihood = -6.48     Transmembrane  60 - 76 ( 58 - 80)
40   INTEGRAL    Likelihood = -6.32     Transmembrane  99 - 115 ( 95 - 122)
      INTEGRAL    Likelihood = -5.10     Transmembrane  435 - 451 ( 432 - 454)
      INTEGRAL    Likelihood = -3.29     Transmembrane  407 - 423 ( 407 - 424)
      INTEGRAL    Likelihood = -2.55     Transmembrane  252 - 268 ( 252 - 268)
      INTEGRAL    Likelihood = -2.02     Transmembrane  131 - 147 ( 130 - 149)
45   INTEGRAL    Likelihood = -1.70     Transmembrane  173 - 189 ( 173 - 189)
      PERIPHERAL    Likelihood = 1.43      21
      modified ALOM score: 2.95

50   *** Reasoning Step: 3
      ----- Final Results -----
      bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
55   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC00276 GB:AF008220 YtgP [Bacillus subtilis]
Identities = 178/545 (32%), Positives = 302/545 (54%), Gaps = 26/545 (4%)

60   Query: 24  QMVKGTAWLTAGNFISRLGAIYIIPWYAWMGKHAEEANALFGMGYEIYALFLLISTVGI 83
      ++++GT  LT G +ISR+LG +Y+IP+  +G  A  ALF  GY  Y  LFL  I+T+G
      Sbjct: 4  KLLRSTFVLTGTYISRILGMVYLIPFSIMVG---ATGGALFQYGYNQYTLFLNIATMGF 60

```

-1326-

Query: 84 PVA VAKQVSKYNTLGKEEMSIYLVVRKILQFMILILGGIFALIMYIGSPLEFASLSKGGQE-- 141
 P AV+K VSKYN+ G E S +++ + ML+ G I I+Y+ +P+FA +S GG++
 Sbjct: 61 PAAVSKFVSKYNSKGDYETSRKMLKAGMSVMLVTGMIAFFILYLSAPMFAEISLGGKDN 120

5 Query: 142 -----LVPILRSLTLAVLVFPSPMSVLRGFFQGFNNLKPYAISQVAEQIIRVIWMLLTAF 195
 +V ++R ++LA+LV P MS++RGFFQG + P A+SQV EQI+R+I++L F
 Sbjct: 121 GLTIDHVYVIRMVSLALLVVPIMSLVRGFFQGHQMMGPTAVSQVVEQIVRIIFLLSATF 180

10 Query: 196 YIMRLGSGDYIAAVTQSTFAAFVGMFASIAVLLYFLW--RYNMLSALIGKTPKHILKDTK 253
 I+++ +G + AV +TFAA +G F + V+LY W R L A++ T L K
 Sbjct: 181 LILKVFNGGLVIAVGYATFAALIGAFGGL-VVLYIYNKRKGSLLAMMENTGTANLSYK 239

15 Query: 254 EILITIKEAIPFIITGAATQIFKLIDQFSFGNTM--ALFTNYSSEELRVMFAYFSSNPG 311
 ++ E A P++ G AI ++ ID +F M A S + L ++ Y
 Sbjct: 240 KMFFELFSYAAPYVFGGLAIPLYNYIDTNTFNKAMIEAGHQAI SQDMLAILTLYVQ--- 295

20 Query: 312 KVTMILIAVATAIAGVGIPLLTENFVKNDKKAARLVNNLQMLLMFLPAVAGSVILAK 371
 K+ MI +++ATA IP +TE+F + K + + +Q +L ++PAV G +L+
 Sbjct: 296 KLVMPVSLATAFGLTLIPTITESFTSGNYKLLNQINQTMQTILFLIIPAVVGISLLSG 355

25 Query: 372 PLYTVFYGL----PQQQALGLFVISLIQTIILSIYTVLAPMLQALFENRKAIIFYLYGLV 427
 P YT FYG P+ A L S + I+ S++TV A +LQ + + + A++ + G+V
 Sbjct: 356 PTYTFYGSSELHPGLGANILLWYSPV-AILFSLFTVNAAILQGINKQKFAVSVLVIGVV 414

30 Query: 428 AKVILQLPSIFLFHAYGPLFSTTVALCIPVILMYLKIHEITGFKRQAIRRTSALVLILTL 487
 K++L +P I L A G + +T + ++ ++ I G+ + + + + L+L+L+
 Sbjct: 415 IKLVNLVPLIKMQADGAILATAGYIASLLYGFIMIKRHAGYSYKILVKRTVLMVLVLSA 474

35 Query: 488 LMSFIISMIWLMNLVI-VPSRLVSLVYIIVIGAIGLVYGFMA LATHLLDKMIGSRAQ 546
 +M + ++ W++ I D ++ + + +++ A+G VY + L K++G R
 Sbjct: 475 IMGIAVKIVQWVLGFFISYQDGMQAATVVVIAAAVGGAVYLYCGYRLGFLQKILGRRRLP 534

Query: 547 DLRRK 551
 RK
 Sbjct: 535 GFFRK 539

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3681> which encodes the amino acid sequence <SEQ ID 3682>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

40	INTEGRAL	Likelihood = -8.60	Transmembrane	468 - 484 (466 - 493)
	INTEGRAL	Likelihood = -8.39	Transmembrane	305 - 321 (299 - 323)
	INTEGRAL	Likelihood = -7.75	Transmembrane	343 - 359 (340 - 362)
	INTEGRAL	Likelihood = -6.58	Transmembrane	374 - 390 (373 - 398)
45	INTEGRAL	Likelihood = -4.25	Transmembrane	138 - 154 (137 - 157)
	INTEGRAL	Likelihood = -3.45	Transmembrane	100 - 116 (98 - 122)
	INTEGRAL	Likelihood = -3.40	Transmembrane	415 - 431 (410 - 432)
	INTEGRAL	Likelihood = -3.35	Transmembrane	499 - 515 (499 - 519)
	INTEGRAL	Likelihood = -2.60	Transmembrane	433 - 449 (432 - 451)
50	INTEGRAL	Likelihood = -2.50	Transmembrane	173 - 189 (173 - 190)
	INTEGRAL	Likelihood = -0.59	Transmembrane	201 - 217 (201 - 220)

----- Final Results -----

55	bacterial membrane	--- Certainty=0.4439(Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC00276 GB:AF008220 YtgP [Bacillus subtilis]
 60 Identities = 169/536 (31%), Positives = 295/536 (54%), Gaps = 24/536 (4%)

Query: 14 MVQGAANSTAGNFISRLGLVLYIIPWYIWMGQYAIQANALFNMGVNVYAYFLLISTTGLN 73
 +++G T G +ISR+LG++Y+IP+ I +G ALF GYN Y FL I+T G
 Sbjct: 5 LLRGFTVLTGLTYISRILGMVYLIPFSIMVGA---TGGALFQYGYNQYTLFLNIATMGFP 61

65

-1327-

Query: 74 VAIKQVAKYNSMGQTEHSYQLIRSTLKLMLGLGLIFSAIMYLGSPFLFASLS-GGDDT-- 130
 A++K V+KYN S G E S +++++ + +ML G+I I+YL +P+FA +S GG D
 Sbjct: 62 AAVSKFVSKYNSKGDYETSRKMLKAGMSVMLVTGMTAFFILYLSAPMPAEISLGGKDNG 121

5 Query: 131 -----LVPIMHSLSLAVFIFPVM SVIRGIFQGHNNIKPYAVSQIAEQIRVIWMLLTTF 185
 +V ++ +SLA+ + P+MS++RG FQGH + P AVSQ+ EQ++R+I++L TF
 Sbjct: 122 LTIDHVYVIRMVSLALLVPMISLVRGFFQGHQMMGPTAVSQVVEQIVRIIFLLSATFL 181

10 Query: 186 IMKLGSGDYASAVTQSTFAAFIGMVASMGVLGYLW--KQGLLAAIFSKPDHTVSDIKG 243
 I+K+ +G AV +TFAA IG + VL Y W ++G L A+ T ++ K
 Sbjct: 182 ILKVFNGGLVIAVG YATFAALIGAFGGLVVL--YIWNKRKGSLLAMPNTGPTANLSYKK 240

15 Query: 244 LLETLKESIPFIVTGSIAIQALIDQWTFVNTMTLFTDYSRSQ--LLVLFYGFNANPAK 301
 + E + P++ G AI + ID TF M + SQ L +L Y K
 Sbjct: 241 MFFELFSYAAPYV FVGLAIPLYNYIDTNTFNKAMIEAGHQAI SQDMLAILTLYVQ---K 296

20 Query: 302 ITMVLIAVAASIGGVGIALLTENYVKKDKMAARLIINNIEMLMFLPALTGAILARP 361
 + M+ +++A + G I +TE++ + K + I ++ ++ ++PA+ G +L+ P
 Sbjct: 297 LVMIPVSLATAFGLTLIPTITESFTSGNYKLLNQINQTMQTILFLIIPAVVGISLLSGP 356

25 Query: 362 LYSVFGASE---ERAIHLFVAVLFQTLALLALYTLFSPMLQALFENRKAIYYFAYGILIK 418
 Y+ FYG+ E ++ + +L +L+T+ + +LQ + + + A+ G++IK
 Sbjct: 357 TYTFYFGESESLHPELGANILLWYSPVAILFSLFTVNAILQGINQKFAVVSIVIGVVIK 416

30 Query: 419 LVLQIPLIYLLHAYGPLLATTIALVVPYIYLMYRRLYQVTHFNKLLQKRLLLTLIETLLM 478
 LVL +PLI L+ A G +LAT + + + + + ++ K+L KR +L L+ + +M
 Sbjct: 417 LVLNVPLIKMQADGAILATAGLYIASLLYGFIMIKRHAGYSYKILVKRTVLMVLVLSAIM 476

35 Query: 479 GLVVFVANWLLGYAFK-PTGRLTSLLYLLIIGGLGMTVYTALTLLTHQLDKLIGSK 533
 G+ V + W+LG+ G++ + + ++I +G VY L K++G +
 Sbjct: 477 GIAVKIVQWVLGFFISYQDGMQAAIVVVIAA AVGGA VYLYCGYRLGFLQKILGRR 532

An alignment of the GAS and GBS proteins is shown below.

Identities = 320/541 (59%), Positives = 431/541 (79%)

35 Query: 12 MSQKTTKVSQQEQMVKGTAWLTAGNFISRLLGAIYIIPWYAWMGKHAEEANALFGMGYEI 71
 MS + +++Q+E MV+G AW TAGNFISRLLG +YIIPWY WMG++A +ANALF MGY +
 Sbjct: 1 MSTEEKQLTQEELMVQGAAWSTAGNFISRLLGVLVIIPWYIWMGQYAIQANALFNMGYNV 60

40 Query: 72 YALFLLISTVGIPVAVAKQVSKYNTLGKEMSIYLVKILQFMLILGGIFALIMYIGSPL 131
 YA FLLIST G+ VA+AKQV+KYN++G+ E S L+R L+ ML LG IF+ IMY+GSPL
 Sbjct: 61 YAYFLLISTTGLNVIAIAKQVAKYNSMGQTEHSYQLIRSTLKLMLGLGLIFSAIMYLGSP 120

45 Query: 132 FASLSKGGQELVPIRLSLTLAVLVFPMSVLRGFFQGFNNLKPYAISQVAEQIIRVIWML 191
 FASLS G LVPI+ SL+LAV +FP MSV+RG FQG NN+KPYA+SQ+AEQ+IRVIWML
 Sbjct: 121 FASLSGGDDTLVPIMHSLSLAVFIFPVM SVIRGIFQGHNNIKPYAVSQIAEQIRVIWML 180

50 Query: 192 LTAFYIMRLGSGDYIAAVTQSTFAAFVGMFASIAVLLYFLWRYNMLSALIGKTPKHILKD 251
 LT F+IM+LGSGDY +AVTQSTFAAF+GM AS+ VL Y+LW+ +L+A+ K + +D
 Sbjct: 181 LTTFFIMKLGSGDYASAVTQSTFAAFIGMVASMGVLGYLWKQGLLAAIFSKPDHTVSD 240

55 Query: 252 TKEILTIETIKEAIPFIITGAAIQIFKLIDQFSFGNTMALFTNYSSEELRVMPAYFSSNPG 311
 K +L+ET+KE+IPFI+TG+AIQ F+LIDQ++F NTM LFT+YS +L V+F YF++NP
 Sbjct: 241 IKGLLLETLKESIPFIVTGSIAIQALIDQWTFVNTMTLFTDYSRSQLLVLFYGFNANPA 300

60 Query: 312 KVTMILIAVATAIAGVGIPLLTENFVKNDKKAARLVVNNLQMLLMFLLPVAVGSVILAK 371
 K+TM+LIAVA +I GVGI LLTEN+VK D KAAARL++NN++ML+MFLLEPA+ G++ILA+
 Sbjct: 301 KITMVLIAVAASIGGVGIALLTENYVKKDKMAARLIINNIEMLMFLPALTGAILARP 360

65 Query: 372 PLYTVFYGLPQGGALGLFVISLIQTHLSIYTVLAPMLQALFENRKAIYFYLGLVAKVI 431
 PLY+VFG + +A+ LFV L QT++L++YT+ +PMLQALFENRKAI YF YG++ K++
 Sbjct: 361 PLYSVFYGASEERAIHLFVAVLFQTLALLALYTLFSPMLQALFENRKAIYFAYGILIKLV 420

Query: 432 LQLPSIFLFHAYGPLFSTTVALCIPVILMYLKIHEITGFKRQAIRRTSALVLILTLLMSF 491
 LQ+P I+L HAYGPL +TT+AL +P+ LMY +++++T F R+ +++ L LI TLLM
 Sbjct: 421 LQIPLIYLLHAYGPLLATTIALVVPYIYLMYRRLYQVTHFNKLLQKRLLLTLIETLLMGL 480

-1328-

Query: 492 IISMIIWLMNLVIVPDSRLVSLVYIIVIGAIGLVYGFMALATHLLDKMIGSRAQDLRRKL 552
 ++ + WL+ P RL SL+Y+++IG +G+ VY + L TH LDK+IGS+A LR+KL
 Sbjct: 481 VVFVANWLLGYAFKPTGRRLTSLLYLLIIGGLGMTVYTALTLTHQLDKLIGSKASRLRQKL 541

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1183

A DNA sequence (GBSx1259) was identified in *S. agalactiae* <SEQ ID 3683> which encodes the amino acid sequence <SEQ ID 3684>. Analysis of this protein sequence reveals the following:

10 Possible site: 40
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 15 bacterial cytoplasm --- Certainty=0.4104(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:BA06290 GB:AP001515 UDP-N-acetylmuramoylalanyl-D-glutamyl-2,
 6-diaminopimelate ligase [Bacillus halodurans]
 Identities = 153/468 (32%), Positives = 237/468 (49%), Gaps = 23/468 (4%)

Query: 33 NVTFNALSYDSRQISSDTLFFA-KGATFK-KEYLDSAITAGLSFYVSETDYGADIPVILV 90
 N +++ DSR++ LFF KG T +Y A++ G VSE +PV++V
 25 Sbjct: 21 NPDHSHHMSREVVGGGLFFCIKGYTVDGHDYAQQAVSNGAVAVVSEPLELSVPV 80

Query: 91 NDIKKAMSLISMSFYNNPQNKLLAFTGKTGKTTAAYFAYHMLKVNHR-PAMLSTMNTT 149
 D ++AM+ ++ FY P N L+L+ TGT GKTT + +++ + ++ TM T
 30 Sbjct: 81 RDSRRAMAQVATKFYGEPTNDLQLIGVTGTNGKTTTITHLIEKIMQDQKMTGLIGTMYTK 140

Query: 150 LDGKSFFKSHLTTPESLDLFRMMATAVENQMTHLIMEVSSQAYLTKRVYGLTFDVGVFILN 209
 + G ++ TTPESL L R A ++ +T +MEVSS A + RV G FDV VF N
 35 Sbjct: 141 I-GHEHKETKNTTPESLVLRQTFADMKKSGVTAMMEVSSHALQSGRVRGCDPDAVAVFSN 199

Query: 210 ISPDHIGPIEHPTFEDYFFHKRLIME-----NSNAVNVN----SQMDHFNIVKEQVEYI 259
 ++PDH+ H T E Y F K LL V+N + D + QV
 40 Sbjct: 200 LTPDHLDD--YHGTMEYKFAKGLLFAQLGNTYQGVAVLNADDPASADFAEMTIAQVVTY 257

Query: 260 PHDFYGDY-SENVITESKAFSPHVKGLLEN-TYDIKLIKFNQENAIAGLACLRGVS I 317
 + D+ +ENV S +F+ E I LIGKF+ N +AA A GV +
 45 Sbjct: 258 GIENEADFQAENVRITSTGTFELAAFEERMELSIHLIGKFSVYNVLA AAAAYVSGVPL 317

Query: 318 EDIKNGIAQTT-VPGRMEVLTQTNGAKIFVDYAHNGDSLKLLAVVEEHQKGDIIILVIGA 376
 ++IK + + V GR E + + VDYAH DSL+ +L V E KGD+ +V+G
 50 Sbjct: 318 QEIKKSLEEVKGVAGRFETVKHDQPFTVIVDYAHTPDSLENVLKTVGELAKGDVRVVVGC 377

Query: 377 PGKKGQSRKDFGDVINQHFNQVILTADDPNFDPLVISQEIASHINRPVTIII-DREE 435
 G++ +++R ++ N Q I T+D+P E+P+ I +++ ++I DR+E
 55 Sbjct: 378 GGDRDKTKRPVMAEIAATTFAN-QAIFTSNPRSEEPMDILRDMEQGAKGDSYLMIEDRKE 436

Query: 436 AIANASTLTNCKLDATIIAGKGADAYQIIKGNRDNYSGLLEVAKKYLK 483
 AI A L + D I+IAGKG + YQ + ++ D VA++ +K
 60 Sbjct: 437 AIFKAIEIAK-EDDIIVIAKGHETYQQFRDRTIDFD-DRIVAQQAIK 482

- 55 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3685> which encodes the amino acid sequence <SEQ ID 3686>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----

-1329-

bacterial cytoplasm --- Certainty=0.4717(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 350/482 (72%), Positives = 399/482 (82%), Gaps = 1/482 (0%)

Query: 1 MITIDKILEILKNDHNFREILFHEHYNNWTQNVTFNALSYSRQISSDTLFFAKGATFK 60
 10 Sbjct: 1 MITIEQLLDILKKDHNFRVLDADGYHYHY-QGFSFERLSYSRQVDGKTLFFAKGATFK 59

Query: 61 KEYLDSAITAGLSFYVSETDYGADIPVILVNDIKKAMSLISMSFYNNPQNKLKLLAFTGT 120
 +YL AIT GL Y+SE DY IPV+LV DIKKAMSLI+M+FY NPQ KLKLLAFTGT
 15 Sbjct: 60 ADYLKEAITNGLQLYISEVDYELGIPVVLVTDIKKAMSLIAMAFYGNPQEKLLKLLAFTGT 119

Query: 121 KGKTTAAYFAYHMLKVNHRPAMLSTMNTTLDGKSFFKSHLTTPESLDLFRMMATAVENQM 180
 KGKTTAAYFAYHMLK +++PAM STMNTTLDGK+FFKS LITPESLDLF MMA V N M
 Sbjct: 120 KGKTTAAYFAYHMLKESYKPMFSTMTNTTLDGKTFFKSQLTTPESLDLFAMMAECVTNGM 179

Query: 181 THLIMEVSSQAYLTKRVYGLTFDVGVLNISPDIHIGPIEHPTFEDYFFHKRLLMENSNAV 240
 THLIMEVSSQAYL RVYGLTFDVGVLNISPDIHIGPIEHPTFEDYF+HKRLLMENS AV
 20 Sbjct: 180 THLIMEVSSQAYLVDRVYGLTFDVGVLNISPDIHIGPIEHPTFEDYFVHKRLLMENSRAV 239

Query: 241 VVNSQMDHFNIVKEQVEYIPHDYFGDYSENVITESKAFSFHVKGKLENTYDIKLIGKFNQ 300
 V+NS MDHF+ + +QV H FYG S+N IT S+AFSF KG+L YDI+LIG FNQ
 25 Sbjct: 240 VINSQMDHFSFLADQVADQEHVFYGLSDNQITTSQAFSFEAKGQLAGHYDIQLIGHFNQ 299

Query: 301 ENAIAAGLACLRLGVSIEDIKNGIAQTTVPGRMEVLTQINGAKIFVDYAHNGDSLKLLA 360
 ENA+AAGLACLRLG S+ DI+ GIA+T VPGRMEVLT TN AK+FVDYAHNGDSL+KLL+
 30 Sbjct: 300 ENAMAAGLACLRLGASLADIQKGIKTRVPGRMEVLTMTNHAKVFVDYAHNGDSLKLLS 359

Query: 361 VVEEHQKGDIIILVLGAPGNKGQSRKDFGVDVINQHNPQLVILTADDPNFEDPLVISQEI 420
 VVEEHQ G ++L+LGAPGNKG+SRR DFG VI+QHPNL VILTADDPNFEDP IS+EIA
 35 Sbjct: 360 VVEEHQTKMLLILGAPGNKGESRRADFGRVHQPNTLVILTADDPNFEDPEDISKEIA 419

Query: 421 SHINRPVTIIIDREEAIANASTLTNCKLDALIIAGKGADAYQIIKGNRDNYSGLLEVAKKYL 482
 SHI RPV II DRE+AI A +L DA+IIAGKGADAYQI+KG + Y+GDL +AK YL
 Sbjct: 420 SHIARPVEIISDREQAIQKAMSLCQGAQDAVITAGKGADAYQIVKQQVAYAGDLATAKHYL 481

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1184

A DNA sequence (GBSx1260) was identified in *S.agalactiae* <SEQ ID 3687> which encodes the amino acid sequence <SEQ ID 3688>. Analysis of this protein sequence reveals the following:

45 Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

55 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1330-

Example 1185

A DNA sequence (GBSx1261) was identified in *S.agalactiae* <SEQ ID 3689> which encodes the amino acid sequence <SEQ ID 3690>. This protein is predicted to be FhuA (fepC). Analysis of this protein sequence reveals the following:

```

5   Possible site: 54
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2785(Affirmative) < succ>
10   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9975> which encodes amino acid sequence <SEQ ID 9976> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAF98153 GB:AF251216 FhuC [Staphylococcus aureus]
    Identities = 141/259 (54%), Positives = 193/259 (74%)

    Query: 7  MSHIKAENIIVSYDQKEIINNLSLSILNQKITTTIIGANGCGKSTLLKALTRIHKIKDGTI 66
20   M+ + + + + Y IIN L + I + K+T+IIG NGCGKSTLLKAL+R+ +K+G +
    Sbjct: 1  MNRLHGQQVKIGYGDNTIINKLDVEIPDGKVTSTIIGPNCGKSTLLKALSRLLA VKEGEV 60

    Query: 67  TIDGHDIAHLPTKEIAKKIALLPQVLEATEGITVYELISYGRFPHQKYLGNLTNDDRSKI 126
    +DG +I TKEIAKKIA+LPQ E +G+TV EL+SYGRFPHQK G LT +D+ +I
25   Sbjct: 61  FLDGENIHTQSTKEIAKKIAILPQSPEVADGLTVGELVSYGRFPHQKGFRLTAEDKKEI 120

    Query: 127 HWAMEMTINVAQFANRDVDDLSSGGQRQKVWIAMALAQDITDIFLDEPTTYLDMNHQLEVLE 186
    WAME+T F +R ++DLSSGGQRQ+VWIAMALAQ TD IFLDEPTTYLD+ HQLB+LE
30   Sbjct: 121 DWAMEVTGTDTFRHRSINDLSSGGQRQVWIAMALAQRTDIIFLDEPTTYLDICHQLEILE 180

    Query: 187 LLKKLNDETQKTIIMVLHDLNLSARYSDYLVAMKIGKIIYEGSPSQIMTKDIKIDIFKID 246
    L++KLN E TI+MVLHD+N + R+SD+L+AMK G II GS ++T++I++ +F ID
    Sbjct: 181 LVQKLNQEQGCTIVMVLHDINQAIRFSDHLIAMKEGDIIATGSTEDVLTQEILEKVFNIID 240

35   Query: 247 AHIIQDPISKQPVLLSYQL 265
    + +DP + +P+L++Y L
    Sbjct: 241 VVLSKDPKTGKPLLVTYDL 259

```

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1929> which encodes the amino acid sequence <SEQ ID 1930>. Analysis of this protein sequence reveals the following:

```

    Possible site: 48
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
45   bacterial cytoplasm --- Certainty=0.2970(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

50   Identities = 166/259 (64%), Positives = 208/259 (80%)

    Query: 7  MSHIKAENIIVSYDQKEIINNLSLSILNQKITTTIIGANGCGKSTLLKALTRIHKIKDGTI 66
    M+ I AE++ ++Y+Q+ II+ LS I KITTTIIGANGCGKS+LLKALTR+ K G +
55   Sbjct: 1  MTTISAEDLTIAEYQRTIIDKLSFYIPEGKITTTIIGANGCGKSSLLKALTRLLPPKQGVV 60

    Query: 67  TIDGHDIAHLPTKEIAKKIALLPQVLEATEGITVYELISYGRFPHQKYLGNLTNDDRSKI 126
    ++G +IA L TKE+AKK+ALLPQV EAT GITVYEL+SYGRFPHQ Y GNL+ D+ I
    Sbjct: 61  YLNGQNIATLETKEVAKKLALLPQVQEATNGITVYELVSYGRFPHQSYFGNLSPADKKAI 120

```


-1331-

Query: 127 HWAMEMTNVAQFANRDVDDLGGGQRQKVWIAMALAQDTDTIFLDEPTTYLDMNHQLEVLE 186
 HWAM+ TNV +A++ VD LSGGQRQ+VW+AMALAQ TDTIFLDEPTTYLD+NHQLE+LE
 Sbjct: 121 HWAMQATNVMAQYADQPVDAALSGGQRQVWLMALAQDTDTIFLDEPTTYLDLNHQLLEILE 180

5 Query: 187 LLKKLNDETQKTIIMVLHDLNLSARYSDYLVMKTKGIIEGSPSQIMTKDIIKDIFKID 246
 L+K LN + KTI+MVLHDLNLSARYSD+L+AMK GKI Y G+ + +MT II+DIF+I
 Sbjct: 181 LVKSLNKDAGKTIVMVLHDLNLSARYSDHLIAMKHGKIHYTGTLADVMTPPIIQDIFQIK 240

10 Query: 247 AHIIQDPISKQPVLVLSYQL 265
 ++ DPI P++L+YQL
 Sbjct: 241 PVLVDDPIHNCPVLVLYQL 259

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 1186

A DNA sequence (GBSx1262) was identified in *S.agalactiae* <SEQ ID 3691> which encodes the amino acid sequence <SEQ ID 3692>. This protein is predicted to be ferrichrome ABC transporter. Analysis of this protein sequence reveals the following:

20 Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 25 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07609 GB:AP001520 ferrichrome ABC transporter
 (ferrichrome-binding protein) [Bacillus halodurans]
 30 Identities = 94/301 (31%), Positives = 177/301 (58%), Gaps = 11/301 (3%)
 Query: 6 IIVLTLLTFFLV---SCGQQTQESTKTTISK--MPKIEGFTYYGKIPENPKKVINFTYS 60
 +++LT+L F L+ +CG T E S+ M E T ++P NP++V+
 35 Sbjct: 7 LLLLTMLLFALLVVAACGSNTDAEQADELESEDEGMITYESETGPIEVPANPQRVV--ALG 64
 Query: 61 YTGYYLLKLGNNVSSSYSLDLEKDSPVFGKQLKEAKKLTADDTEAIAAQKPDLMVFDQDPN 120
 +TG +L L VNV K++P + + L++ +++ ++ E I PDLI+ + N
 Sbjct: 65 FTGNILALDVNVGVDT-WSKNNPNYEQLQDVTEVSEENLEQIMELDPDLIIAYSTVQN 123
 40 Query: 121 INTLKKIAPTLLVIKYGAQNYLDMMPALGKVFGEKEANQWVSQWTKTLAVKKDLHHILK 180
 L++IAPT++ Y +YL+ +GK+ KE+EA WV +K + +++ +
 Sbjct: 124 AEQLQEIAPTVLYTYNNLDYLEQHVEIGKLLNKEEBAQAWVDDFKARAEQAGEBBIKEKIG 183
 45 Query: 181 PNTTFTIMDFYDKNIYLYGNNGRGELIYDSLGYAAPEKVKKDVFKGWFTVSQEAIGD 240
 + T ++++ ++ +Y++GNN+GRG E++Y ++ A PE+V++ G++ +S EA+ +
 Sbjct: 184 EDATVSVIETFEDQLYVFGNNWGRGTETILYQIMDLAMPERVEEMALADGYALSFEALPE 243
 Query: 241 YVG DYALVNINKTTKKAASSLKESDVWKNLPAVKKGHIIESNYDVVFYFSDPLSLEAQLKSF 30`
 + GDY +++ N +A +S +E++ ++++PAV+ G + E+N FYF+DPLSLE QL+ F
 50 Sbjct: 244 FAGDYIILSKN---DEADNSFQETINTYQSIPAVQNGQVFEANAKFIFYNDPLSLELQLEFF 301

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3693> which encodes the amino acid sequence <SEQ ID 3694>. Analysis of this protein sequence reveals the following:

55 Possible site: 19
 >>> May be a lipoprotein
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1332-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:BAB07609 GB:AP001520 ferrichrome ABC transporter
 (ferrichrome-binding protein) [Bacillus halodurans]
 Identities = 112/306 (36%), Positives = 178/306 (57%), Gaps = 3/306 (0%)

10 Query: 2 KKLTLTLLTLCITITLIACGNQATNHSNTASKSLSPMPQIAGVTYYGDIPKQPKRVVSLA 61
 K L LL L + + ACG+ +S M T ++P P+RVV+L
 Sbjct: 5 KHLTLLTLLTLLFALLVVAACGSNTDARQADELESEEDGMITYESETGPIEVFANPQRVVVALG 64

15 Query: 62 STYTGYLKKLDMNLVGVTSYDKKNPILAKTVKKAKQVAATDLEAVTTLKPDLLIVVGSTEE 121
 +TG + LD+N+VGV ++ K NP + ++ +V+ +LE + L PDLI+ ST +
 Sbjct: 65 --FTGNILALDVNVVGVDTWSKNNPNYEQLLDQVTEVSEENLEQIMELDPDLIIAYSTVQ 122

20 Query: 122 NIKQLAEIAPVISIEYRKRDLQVLSDFGRIFNKEDKAKKWLKDWKTKTAAYEKEVKAVT 181
 N +QL EIAP + Y DYL+ + G++ NKE++A+ W+ D+K + +E+K
 Sbjct: 123 NAEQLQEIAPTPLYTYNNLDYLEQHEVEIGKLLNKEEBAQAWDDFKARAEQAGEEIKEKI 182

25 Query: 182 GDKATFTTIMGLYEKDVYLFKGDWGRGGEIIHQAFHYDAPEKVKTEVFKQGYLSLSQEVLP 241
 G+ AT +++ +E +Y+FG +WGRG EI++Q PE+V+ GY +LS E LP
 Sbjct: 183 GEDATVSVIETFRDQLYVFGNNWGRGTEILYQTMDLAMPERVEEMALADGYALSFEALP 242

30 Query: 242 DYIGDYVVVAEDDKTGSALYESKLWQSIPAVKKHHVIKVNANVFYFTDPLSLEYQLETL 301
 ++ GDY+++ +++D+ ++ E+ +QSIPAV+ V + NA FYF DPLSLE QLE
 Sbjct: 243 EFAGDYIIL-SKNDEADNSFQETINTYQSIPAVQNGQVFEANAKFVFNDPLSLELQLEFF 301

Query: 302 REAILS 307
 +E LS
 Sbjct: 302 KEHFLS 307

An alignment of the GAS and GBS proteins is shown below.

Identities = 140/316 (44%), Positives = 212/316 (66%), Gaps = 12/316 (3%)

35 Query: 1 MKKIGLIIV-LTLLTFFLVSCGQQTQESTKIT--ISKMPKIEGFTYYGKIPENPKKVINF 57
 MKK+ +++ L L T L++CG Q S + +S MP+I G TYYG IP+ PK+V++
 Sbjct: 1 MKKLTLTLLTLCITITLIACGNQATNHSNTASKSLSPMPQIAGVTYYGDIPKQPKRVVSL 60

40 Query: 58 TYSYTGYYLLKLVN---VSSYSLDLKDSPVFGKQKLEAKKLTAADTEAIAAQKPDLMIV 114
 +YTGYL KL +N V+SY +K +P+ K +K+AK++ A D EA+ KPDLI+V
 Sbjct: 61 ASTYTGYLKKLDMNLVGVTSY---DKKNPILAKTVKKAKQVAATDLEAVTTLKPDLLIV 116

45 Query: 115 FDQDPNINTLKKIAPTLLVIKYGAQNYLDMMPALGKVFGEKEANQWVSQWKTTLAVKKD 174
 + NI L +IAP + I+Y ++YL ++ G++F KE +A +W+ WKTKT A +K+
 Sbjct: 117 GSTEENIKQLAEIAPVISIEYRKRDLQVLSDFGRIFNKEDKAKKWLKDWKTKTAAYEKE 176

50 Query: 175 LHHILKPNTTFTTIMDFYDKNIYLYGNNFGRGGELIYDSLGYAAPEKVKDVFKKGWFTVS 234
 + + TFTIM Y+K++YL+G ++GRGGE+I+ + Y APEKVK +VFK+G+ ++S
 Sbjct: 177 VKAVTGDKATFTTIMGLYEKDVYLFKGDWGRGGEIIHQAFHYDAPEKVKTEVFKQGYLSLS 236

55 Query: 235 QEAIIGDYVGDYALVNINKTTKKAASSLKESDVWKNLPAVKKGHIIESNYDVFFYFSDPLSL 294
 QE + DY+GDY +V K S+L ES +W+++PAVKK H+I+ N +VFYF+DPLSL
 Sbjct: 237 QEVLDPYIGDYVVVAEE--DDKTGSALYESKLWQSIPAVKKHHVIKVNANVFYFTDPLSL 294

Query: 295 EAQLKSFTKAIKENTN 310
 E QL++ +AI + N
 Sbjct: 295 EYQLETLREAILSEN 310

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1333-

Example 1187

A DNA sequence (GBSx1263) was identified in *S.agalactiae* <SEQ ID 3695> which encodes the amino acid sequence <SEQ ID 3696>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3431(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
15 vaccines or diagnostics.

Example 1188

A DNA sequence (GBSx1264) was identified in *S.agalactiae* <SEQ ID 3697> which encodes the amino acid sequence <SEQ ID 3698>. This protein is predicted to be ferrichrome transport permease (permease). Analysis of this protein sequence reveals the following:

```

20  Possible site: 39
    >>> May be a lipoprotein
        INTEGRAL    Likelihood = -12.74    Transmembrane    129 - 145 ( 123 - 150)
        INTEGRAL    Likelihood = -10.67    Transmembrane    248 - 264 ( 240 - 283)
        INTEGRAL    Likelihood = -10.14    Transmembrane    205 - 221 ( 196 - 228)
25  INTEGRAL    Likelihood = -5.95    Transmembrane    319 - 335 ( 317 - 336)
    INTEGRAL    Likelihood = -3.56    Transmembrane    73 - 89 ( 73 - 90)
    INTEGRAL    Likelihood = -3.19    Transmembrane    288 - 304 ( 288 - 304)
    INTEGRAL    Likelihood = -2.76    Transmembrane    266 - 282 ( 265 - 283)
    INTEGRAL    Likelihood = -2.23    Transmembrane    103 - 119 ( 101 - 122)
30  INTEGRAL    Likelihood = -1.01    Transmembrane    158 - 174 ( 158 - 174)

----- Final Results -----
      bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF98154 GB:AF251216 FluB [Staphylococcus aureus]
Identities = 116/313 (37%), Positives = 194/313 (61%), Gaps = 3/313 (0%)
40  Query: 26  ILFLIGCYASLRFGAINFKTSDLITVLKNPLKNSNAQDVIFDIRLPRIIAAILVGAAMSQ 85
      ++ LI + S G + S + I + N ++ Q++I +IR+PR IAA++VG A++
      Sbjct: 28  MILLITLFLSTLIGDAKIQAFTIIEAIFNYNPSNQQQNIINEIRIPRNIAAVIVGMALAV 87

45  Query: 86  AGAIMQGVTRNAIADPGLLGINAGAGLALVVAYAFGLSMHYSTILIVCLLGSVISCLLVF 145
      +GAI+QGVTRN +ADP L+G+N+GA AL + YA L + + ++ LG+++ +V
      Sbjct: 88  SGAI IQGVTRNGLADPALIGLNSGASFALALTYAVLPNTSFLILMFAGFLGAILGGAIVL 147

50  Query: 146  TLSYTKQKGYHQLRLILAGAMISTLFTSVGQVVTLYFKLNRTVIGWQAGGLSQINWKMLI 205
      + +++ G++ +R+ILAGA +S + T++ Q + L F+LN+TV W AGG+S W L
      Sbjct: 148  MIGRSRRDGFNPMRIILAGAAVSAMLTALSQGIALAFRLNQTVTFWTAGGVSGTTWWSHLK 207

55  Query: 206  IIAPIIIILGLLISQLLAHQLTILSLNESVAKALGQKTQLMTAFLLILVFLSASSVALIG 265
      P+I + L I ++ QLTL+L ES+AK LGQ ++ L+I + L+ +VA+ G
      Sbjct: 208  WAIPILGIALFIILTISKQLTILNLGESLAKGLGQNVMTIRGICLIITAMILAGIAVAIAG 267

```

-1334-

Query: 266 TVSFIGLIIPHFIKLFIPKDYRLLLPLIGFSGATFMIWVDLSSRIINPPSETSISSIISI 325
 V+F+GL++PH + I DY +LPL G ++ D+ +R + E + +IIS
 Sbjct: 268 QVAFVGLMVPFIARFLIGTDYAKILPLTALLGGILVLVADVIARYL---GEAPVGATISF 324

5 Query: 326 VGLPCFLWLIRKG 338
 +G+P FL+L++KG
 Sbjct: 325 IGVFYFLYLKKG 337

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3699> which encodes the amino acid
 10 sequence <SEQ ID 3700>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.09	Transmembrane	256 - 272 (248 - 287)
INTEGRAL	Likelihood = -10.67	Transmembrane	26 - 42 (23 - 48)
15 INTEGRAL	Likelihood = -6.90	Transmembrane	137 - 153 (133 - 157)
INTEGRAL	Likelihood = -5.10	Transmembrane	167 - 183 (166 - 187)
INTEGRAL	Likelihood = -4.57	Transmembrane	213 - 229 (210 - 232)
INTEGRAL	Likelihood = -2.02	Transmembrane	112 - 128 (110 - 131)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.5437(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:AAF98154 GB:AF251216 FhuB [Staphylococcus aureus]
 Identities = 99/274 (36%), Positives = 159/274 (57%), Gaps = 1/274 (0%)

30 Query: 34 LSFSLCVAIYCHLRFGAVALSHQDLNSILFG-KQNGHKANVLLAIRLPRLFGATLTGSAL 92
 LS L + ++ G + + +F + + N++ IR+PR A + G AL
 Sbjct: 26 LSMILLITLFIISTLIGDAKIQAFTIIEAIFNYPNSNQQQNIINEIRIPRNIAAIVIGMAL 85

35 Query: 93 AVSGTIMQAITRNPIAEPGLLGINAGAGLALVLAFAFVPHLHYSLIILLSSLLGSSLAATL 152
 AVSG I+Q +TRN +A+P L+G+N+GA AL L YA +P+ + +++ LG+ L +
 Sbjct: 86 AVSGAIIQGVTRNGLADPALIGLNSGASFALALTYAVLPNTSFLILMFAGFLGAILGGAI 145

40 Query: 153 VFGLSYQSGKGYHQLRLVLGAMVSILLSALGQGITNYYHLANAVIGWQAGGLVGVNWQM 212
 V + G++ +R++LAGA VS +L+AL QGI + L V W AGG+ G W
 Sbjct: 146 VLMIGRSRRDGFNPMRIILAGAAVSAMLTALSQGIALLAFRLNQTVTFWTAGGVSGTTWSH 205

45 Query: 213 IGYIAPLIILSLCLAQLLSYHLTVLSLSSESQAKALGQKTNLISAVFMILVLILSSAAVAI 272
 + + PLI ++L + +S LT+L+L ES AK LGQ +I + +I+ +IL+ AVAI
 Sbjct: 206 LKWAIPILGIALFIILTISKQLTILNLGESLAKGLQNVMTIRGICLIAMILAGIAVAI 265

Query: 273 AGSISFIGLVIPHLMKHFTPHHYRYLLPLCAVSG 306
 AG ++F+GL++PH+ + Y +LPL A+ G
 Sbjct: 266 AGQVAFVGLMVPFIARFLIGTDYAKILPLTALLG 299

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 158/295 (53%), Positives = 214/295 (71%), Gaps = 1/295 (0%)

Query: 6 KKL VQKNKSNHFWLVFFITLILFLIGCYASLRFGAINFKTSDLIITVLKNPLKNSNAQDVI 65
 KK KS+ FWLVF + + Y LRFGA+ DL ++L +N + +V+
 Sbjct: 16 KKTQIITKSHIFWLVFVLLSFLCVAIYCHLRFGAVALSHQDLNSILFGK-QNGHKANVL 74

55 Query: 66 FDIRLPRIIAAILVGAAMSQAGAIMQGVTRNAIADPGLLGINAGAGLALVVAYAFGLSMH 125
 IRLPR+ A L G+A++ +G IMQ +TRN IA+PGLLGINAGAGLALV+AYAF+ +H
 Sbjct: 75 LAIRLPRLFGATLTGSALAVSGTIMQAITRNPIAEPGLLGINAGAGLALVLAFAFVPHLH 134

60 Query: 126 YSTILIVCLLGSVISCLLVFTLSYTKQKGYHQLRLILAGAMISTLFTSVGQVVTLTYFKLN 185
 YS I+++ LLGS ++ LVF LSY KGYHQLRL+LAGAM+S L +++GQ +T Y+ L
 Sbjct: 135 YSLIILLSSLLGSSLAATLVFGLSYQSGKGYHQLRLVLGAMVSILLSALGQGITNYYHLA 194

Query: 186 RTVIGWQAGGLSQINWKMLIIIIPIIIILGLLISQLLAHQLTILSLNESVAKALGQKTQLM 245

-1335-

VIGWQAGGL +NW+M+ IAP+IIL L ++QLL++ LT+LSL+ES AKALGQKT L+
 Sbjct: 195 NAVIGWQAGGLVGVNWQMIGYIAPLIILSLCLAQLLSYHLTVLSLSESAKALGQKTNLI 254

Query: 246 TAFLLLLIVLFLSASSVALIGTVSFIGLIIPHFIKLFIPKDYRLLLPLIGFSGATF 300
 +A +++VL LS+++VA+ G++SFIGL+IPH +K F P YR LLPL SGA+F
 Sbjct: 255 SAVFMILVLILSSAAVAIAGSISFIGLVIPHLMKHFTPHHYRYLLPLCAVSGASF 309

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 1189

A DNA sequence (GBSx1265) was identified in *S.galactiae* <SEQ ID 3701> which encodes the amino acid sequence <SEQ ID 3702>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1492(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 1190

A DNA sequence (GBSx1266) was identified in *S.galactiae* <SEQ ID 3703> which encodes the amino acid sequence <SEQ ID 3704>. This protein is predicted to be ferrichrome transport permease (permease). Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -10.35	Transmembrane	282 - 298 (279 - 309)
INTEGRAL	Likelihood = -7.06	Transmembrane	120 - 136 (115 - 141)
INTEGRAL	Likelihood = -7.01	Transmembrane	62 - 78 (61 - 80)
INTEGRAL	Likelihood = -6.10	Transmembrane	250 - 266 (241 - 272)
INTEGRAL	Likelihood = -5.52	Transmembrane	196 - 212 (190 - 215)
INTEGRAL	Likelihood = -5.47	Transmembrane	155 - 171 (151 - 174)
INTEGRAL	Likelihood = -4.99	Transmembrane	304 - 320 (303 - 322)
INTEGRAL	Likelihood = -3.35	Transmembrane	91 - 107 (90 - 110)

----- Final Results -----
 bacterial membrane --- Certainty=0.5140(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF98155 GB:AF251216 PhuG [Staphylococcus aureus]
 Identities = 122/334 (36%), Positives = 208/334 (61%), Gaps = 3/334 (0%)

Query: 1 MIQKNKAPFVLISSVIILLILV---SISLGYANTSVIDVLKLSGKSDDAFLFIITNI 57
 MI N LI+ + +LL L SI+ G N V K + G+ D I+ +
 Sbjct: 1 MISSNNKRRQLIALAVFSILLFLGCTWSITSGEYNIPVERFFKTLIGQGAIDELILLDF 60

Query: 58 RLPRIIVCIFGCASLGIAGLLQLTKNPLADSGILGINAGAGLVIALTIGTFNVSNPETI 117
 RLPR+++ I GA+L I+G ++Q++TKNP+A+ GILGINAG G IAL I ++

-1336-

Sbjct: 61 RLPRMMITILAGAALSISGAIVQSVTKNPIAEPGILGINAGGGFAIALFIAIGKINADNF 120

Query: 118 LYFLPLFAMFGLVTIFLIYLSYRRNHNISPTRLIVTGIGISTIISGVMILIISQSNQ 177
 +Y LPL ++ GG+ T +I++ S+ +N ++P +++ G+G+ T + G I I+S+ +++

5 Sbjct: 121 VYVLPLISILGGITTALIIIFISFNKNEGVTPASMLVIGVGLQTALYCGSITIMSKFDDK 180

Query: 178 KMDMIVEWLSGKITISSWTTIITFIPILILLWGLAYSRSRHLNIMNLNEQTALALGLHLK 237
 + D I W +G I W +I F+P +++ +S LNI++ + A LG+ L

10 Sbjct: 181 QSDFLAAWFAGNIWDEWPFVIAFLPWVLIITPYLLFKSNTLNIIHTGDNIARGLGVRLS 240

Query: 238 KERIYTLMLTSSLAAISVVLIGNITFIGLLAGHLRRLGNHKKIILPSCLLIGAIILLV 297
 +ER+ + L++ +V + G+I+FIGL+ H++R++G H++ LP +L+GA +L++

Sbjct: 241 RERLILFFIAVLSAAVAVAGSISFIGLMGPHIAKRIVGPRHQLFLPIAILVGCALLVI 300

15 Query: 298 SDTIGRLLLVGTGIPTGLVVSIIIGAPYFLWLMTK 331
 +DTIG+++L G+P G+VV+IIGAPYFL+LM K

Sbjct: 301 ADTIGKIVLPQGGVPAGIVVAIIGAPYFLYLMYK 334

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1939> which encodes the amino acid sequence <SEQ ID 1940>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -10.93 Transmembrane 254 - 270 (252 - 284)
 INTEGRAL Likelihood = -10.46 Transmembrane 294 - 310 (292 - 320)
 INTEGRAL Likelihood = -6.74 Transmembrane 25 - 41 (18 - 43)
 INTEGRAL Likelihood = -6.26 Transmembrane 103 - 119 (102 - 125)
 INTEGRAL Likelihood = -3.66 Transmembrane 164 - 180 (164 - 186)
 INTEGRAL Likelihood = -3.03 Transmembrane 209 - 225 (207 - 226)
 30 INTEGRAL Likelihood = -2.71 Transmembrane 74 - 90 (74 - 91)
 INTEGRAL Likelihood = -2.13 Transmembrane 326 - 342 (325 - 343)
 INTEGRAL Likelihood = -1.97 Transmembrane 135 - 151 (135 - 151)

----- Final Results -----

35 bacterial membrane --- Certainty=0.5373(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 153/322 (47%), Positives = 229/322 (70%), Gaps = 1/322 (0%)

40 Query: 11 LISSVIIILLIIL-VSISLGyantSVIDLKLSGKSDDAFLFIITNIRLPRIIVCIFGG 69
 L +S+I+LL+ ++ ++SLG ++ S +D++ + GKS A FI+ NIRLPRI+ GG

Sbjct: 22 LYTSILILLVSLMGLALSICESHLSFLDLVHVFLGKSSHAI SFIVINIRLPRI LAACLG 81

45 Query: 70 ASLGIAGLLQLTKNPLADSGILGINAGAGLVIALTIGTFNVSNPTILYFLPLFAMFEG 129
 SL ++GLLQ LT+NPLADSG+LGI GAG+ +A+ + I ++LPLFAM G

Sbjct: 82 GSLALSGLLQLRLTRNPLADSGVLGITIGAGISLAIVVSFSFFEQA HISHYLPFLFAMLGA 141

50 Query: 130 LVTIFLIYLSYRRNHNISPTRLIVTGIGISTIISGVMILIISQSNQKMDMIVEWLSGK 189
 +VT F +Y +S + I PTRLI+TG+ ++T++S +M+ ++ N K+D+++ WLSG+

Sbjct: 142 IVTTFSVYVWLSLTKQGQIDPTRLILTGVAVTTMLSSLMVALVGHINRYKVDLVINWLSGQ 201

55 Query: 190 ITISSWTTIITFIPILILLWGLAYSRSRHLNIMNLNEQTALALGLHLKKERIYTLMLTSS 249
 + W T+ P+L+ W L YS++ LNIM L + TA+ LGL L ++R L+L +

Sbjct: 202 LIGDDWPTLSVIAPLLLCFWLLTYSQAHFLNIMGLADNTAIGLGLPINRKRRLIILVLAAG 261

60 Query: 250 LAAISVVLIGNITFIGLLAGHLRRLGNHKKIILPSCLLIGAIILLVSDTIGRLLLVGT 309
 L A+SV+L+GNI+FIGL+AGH S L+G+NHKI +P +LIG I+LLV+DT+GR+ LVG+

Sbjct: 262 LGALSVLVGNISFIGLIAGHFSTYLVGSNHKTIPTISILIGMILLVADTVGRVYLVGS 321

Query: 310 GIPTGLVVSIIIGAPYFLWLMTK 331
 I TG++VS+IGAPYFL+LM K

Sbjct: 322 NIQTGILVSLIGAPYFLYLMK 343

65 There is also homology to SEQ ID 396.

-1337-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1191

A DNA sequence (GBSx1267) was identified in *S.agalactiae* <SEQ ID 3705> which encodes the amino acid sequence <SEQ ID 3706>. Analysis of this protein sequence reveals the following:

Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3785(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC05779 GB:AF051356 unknown [Streptococcus mutans]
 Identities = 49/93 (52%), Positives = 63/93 (67%)

Query: 1 MILTFNPGKLERQEFFKELINYLWIHDDVTLRKIKSHFTDYSKIDRLLEEYINHG YILRQ 60
 MI +N KL RQ FF +LINYL IHDDVTLR+IK +F D ++R +E+Y+ GY+LR+
20 Sbjct: 1 MIKIYNGDKLTRQPPFFIKLINYLQIHDDVTLRQIKRNFADTEHLERSIEDYVQAGYVLRE 60

Query: 61 NKRYSLNLPFLSSLDGLVLDLDFIDSDS QIYQ 93
 NK Y L +LDGL LD +F+D S IYQ
25 Sbjct: 61 NKHYNAFELLENLDGLTLD S QIFVDDQSSIYQ 93

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3707> which encodes the amino acid sequence <SEQ ID 3708>. Analysis of this protein sequence reveals the following:

Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3447(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 108/212 (50%), Positives = 143/212 (66%)

40 Query: 1 MILTFNPGKLERQEFFKELINYLWIHDDVTLRKIKSHFTDYSKIDRLLEEYINHG YILRQ 60
 MI F+ KL RQ FF++LINYL HD V LR+IK F + + ID+ +E Y+ GYI R+
Sbjct: 1 MITVFHSDKLTRQPPFFQDLINYLQHDHVLREIKKAFENVGTGIDKAIESYVQAGYIRRE 60

Query: 61 NKRYSLNLPFLSSLDGLVLDLDFIDSDS QIYQLLQKRKFVTNLDNPTNHLVFEETDPE 120
 NKRY +NLP +SS L LD ++F+D+ S +Y+ + F T L N TN ++ E+T+
45 Sbjct: 61 NKRYGINLPVSSDQQLALDTMLFVDTC SAMYENILAVVFETQLTNQTNRVMIKEKTINIT 120

Query: 121 RNTLTLSNYFYKLTNGYPLSREQKKLYQLLGDVNSEYALKYMSSFILKFLRKDSVKQKRT 180
 R+ LTL+NYFY+L G S EQ LY LLGDVN EYALKYM++F+LKF RKD V QKR
50 Sbjct: 121 RDDLTLANIFYRLKRGEKPSAEQMDLYDLLGDVNQEYALKYMTTFLKFTRKDFVMQKRP 180

Query: 181 VIFTQALELLGYISLNQDTTYRLNAKLDVEAL 212
 IF++AL LGY+ + TTY+L LD E+L
Sbjct: 181 DIFVEALVTLGYLKQVEPTTYQLLMTLDKESL 212

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1338-

Example 1192

A DNA sequence (GBSx1268) was identified in *S.galactiae* <SEQ ID 3709> which encodes the amino acid sequence <SEQ ID 3710>. Analysis of this protein sequence reveals the following:

```

Possible site: 24
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0824(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAB39104 GB:U57759 intrageneric coaggregation-relevant adhesin
[Streptococcus gordonii]
15 Identities = 261/311 (83%), Positives = 283/311 (90%)

Query: 1  MSKILVFGHQNPDSDAIGSSVAFAYLAKEAWGLDTEAVALGTPNEETAYVLDYFGVQAPR 60
      MSKILVFGHQNPDSDAIGSS AFAYLA+EA+GLDTEAVALG PNEETA+VLDYFGV APR
Sbjct: 1  MSKILVFGHQNPDSDAIGSSVAFAYLAREAYGLDTEAVALGEPNEETAFLDYFGVAAPR 60

20 Query: 61  VVESAKAEGVETVILTDHNEFQQSISDIKDVTYGVVDHHRVANFETANPLYMRLEPVGS 120
      V+ SAKAEG E VILTDHNEFQQS++DI +V VYGVVDHHRVANFETANPLYMRLEPVGS
Sbjct: 61  VITSAKAEGAEQVILTDHNEFQQSVADIAEVEVYGVVDHHRVANFETANPLYMRLEPVGS 120

25 Query: 121 ASSIVYRMPKENGVSVPKELAGLLSGLISDTLLKSPPTTHASDIPVAKELAEAGVNLE 180
      ASSIVYRMPKE+ V+V KE+AGL+LSGLISDTLLKSPPTTH +D +A ELAEAGVNLE
Sbjct: 121 ASSIVYRMPKEHSVAVSKEIAGLMLSGGLISDTLLKSPPTHTDKAIAPELAEAGVNLE 180

30 Query: 181 EYGLEMLKAGTNLSSKTAELIDIDAKTFELNGEAVRVAQVNTVDINDILARQEEIEVAI 240
      EYGL MLKAGTNL+SK+A ELIDIDAKTFELNG VRVAQVNTVDI ++L RQ EIE AI
Sbjct: 181 EYGLAMLKAGTNLASKSAEELIDIDAKTFELNGNNVRVAQVNTVDIAEVLERQAEIEAAI 240

35 Query: 241 QEAIIVTEGYSDFLVIMITDIVNSNSEILA+GSNMKVEAAFEFTLENNHAFLAGAVSRKKQ 300
      ++AI GYSDFVLMITDI+NSNSEILA+GSNM KVEAAF F LENNHAFLAGAVSRKKQ
Sbjct: 241 EKAIADNGYSDFVLMITDIINSNSEILAIGSNMDKVEAAFNFLVLENNHAFLAGAVSRKKQ 300

Query: 301 VVPQLTESYNA 311
      VVPQLTES+NA
Sbjct: 301 VVPQLTESFNA 311
40

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3711> which encodes the amino acid sequence <SEQ ID 3712>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
45 >>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -2.02 Transmembrane 141 - 157 ( 141 - 157)

----- Final Results -----
      bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9103> which encodes the amino acid sequence <SEQ ID 9104>. Analysis of this protein sequence reveals the following:

```

Possible site: 50
55 >>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -2.02 Transmembrane 139 - 155 ( 139 - 155)

----- Final Results -----
      bacterial membrane --- Certainty= 0.181(Affirmative) < succ>
      bacterial outside --- Certainty= 0.000(Not Clear) < succ>
60     bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

```


-1339-

An alignment of the GAS and GBS proteins is shown below.

Identities = 253/311 (81%), Positives = 283/311 (90%)

```

5   Query: 1   MSKILVFQHNPDSDAIGSSVAFAYLAKEAWGLDTEAVALGTPNEETAYVLDYFGVQAPR 60
      MSKILVFQHNPD+DAI SS AF YL+++A+GLDTE VALGTPNEETA+ LDYFGV+APR
      Sbjct: 3   MSKILVFQHNPDDAIASSYAFDYLSQLAFGLDTEVVALGTPNEETAFAFDYFGVEAPR 62

10  Query: 61  VVESAKAEGVETVILTDHNEFQQSISDIKDVTYGVVDHHRVANFETANPLYMRLEPVGS 120
      VVESAKA+G E VILTDHNEFQQSI+DI++V VYGVVDHHRVANFETANPLYMR+BPVGS
      Sbjct: 63  VVESAKAQGSEQVILTDHNEFQQSIADIREVEVYGVVDHHRVANFETANPLYMRVBPVGS 122

      Query: 121 ASSIVYRMFKENGVSVPKELAGLLSGLISDTLLLSKSPTHASDIPVAKELAEAGVNLE 180
      ASSIVYRMFKENG+ VPK +AG+LLSGLISDTLLLSKSPTH SD VA+BLAELA VNLE
15  Sbjct: 123 ASSIVYRMFKENGIEVPKAIAGMLLSGLISDTLLLSKSPTHVSDHLVAEELAEAEVNLE 182

      Query: 181 EYGLEMLKAGTNLSSKTAELIDIDAKTFELNGEAVRVAQVNTVDINDILARQEEIEVAI 240
      +YG+ +LKAGTNL+SK+ ELI IDAKTFELNG AVRVAQVNTVDI ++L RQE IE AI
      Sbjct: 183 DYGMALLKAGTNLASKSEVELIGIDAKTFELNGNAVRVAQVNTVDIAEVLERQEAIEAAI 242

20  Query: 241 QEAIIVTEGYSDFVLMITDIVNSNSEILALGSNMAKVEAAFEFTLENNHAFLAGAVSRKKQ 300
      ++A+ EGYSDFVLMITDIVNSNSEILA+G+NM KVEAAF FTL+NNHAFLAGAVSRKKQ
      Sbjct: 243 KDMAAEGYSDFVLMITDIVNSNSEILAIGANMDKVEAFNFTLDNNHAFLAGAVSRKKQ 302

25  Query: 301 VVPQLTESYNA 311
      VVPQLTES+ A
      Sbjct: 303 VVPQLTESFGA 313

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1193

A DNA sequence (GBSx1269) was identified in *S.agalactiae* <SEQ ID 3713> which encodes the amino acid sequence <SEQ ID 3714>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 20
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2769(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
40  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC05773 GB:AF051356 pyruvate-formate lyase activating enzyme
  [Streptococcus mutans]
45  Identities = 184/260 (70%), Positives = 217/260 (82%)

      Query: 3   EIDYKVTGMIHSTESFGSVDGPGIRFIIFMQGCKMRCQYCHNPDTWEMETNNSKERTVE 62
      ++DY+KVTG+++STESFGSVDGPGIRF++FMQGC+MRCQYCHNPDTW M+ + + ERT
      Sbjct: 4   KVDYEKVTGLVNSTESFGSVDGPGIRFVVFVMQGCQMRCQYCHNPDTWAMKNDRATERTAG 63

50  Query: 63   DVLKEALRYKHFWGKDGGITVSGGEAMLQIDFITALFIEAKKLGIIHTLDTGCFAYRATP 122
      DV KEALR+K FWG GGITVSGGEA LQ+DF+ ALF AK+ GIHTLDTC +R TP
      Sbjct: 64   DVFKALRFKDFWGDGTGKITVSGGEATLQMDFLIALFSLAKEKGIHTLDTCALTFRNTP 123

55  Query: 123  EYHAILEKLDDVTDLVLLDLKEIDSEQHKIVTRQSNKNILQFARYLSDRGTPVWIRHVLV 182
      +Y EKL+ VTDLVLLD+KEI+ +QHKIVT SNK IL ARYLS D G PVWIRHVLV
      Sbjct: 124 KYLEKYEKLMAVTDLVLLDIKEINPDQHKIVTGHSNKTIACARYLSDIGKPVWIRHVLV 183

      Query: 183 PGLTDIDDHLKRLGEFVQTLDNVDKFEVLPYHTMGFEKWRELGIPIPLAGVKPPTPERVK 242
      PGLTD D+ L +LGE+V+TL NV +FE+LPYHTMGFEKWRELGIPIPL GVKPPTP+RV+
60  Sbjct: 184 PGLTDRDEDLIKLGYYVTKLKNVQRFELPYHTMGFEKWRELGIPIPLEGVKPTPDRVR 243

```

-1340-

Query: 243 NAKDIMKTESYTEYLKRIQN 262
 NAK +M TE+Y EY KRI +
 Sbjct: 244 NAKKLMHTETVEYKKRINH 263

5

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3715> which encodes the amino acid sequence <SEQ ID 3716>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4614(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

10

15

An alignment of the GAS and GBS proteins is shown below.

Identities = 223/260 (85%), Positives = 239/260 (91%)

20

Query: 1 MAEIDYKKVTGMIHSTESFGSVDGPGIRFIIFMQGCKMRCQYCHNPDTWEMETNNSKERT 60
 M E DY +VTGM+HSTESFGSVDGPGIRFIIF+QGCK+RCQYCHNPDTWEMETNNSK RT
 Sbjct: 25 MTEKDYGQVTGMVHSTESFGSVDGPGIRFIIFLQGCKLRCQYCHNPDTWEMETNNSKIRT 84

25

Query: 61 VEDVLKEALRYKHFHWGKGGITVSGGEAMLQIDFITALFIEAKKLGIIHTLDTGCFAYRA 120
 V DVLKEAL+YKHFHWGK GGITVSGGEAMLQIDFITALFIEAKKLGIIHTLDTGCF YR
 Sbjct: 85 VNDVLKEALQYKHFHWGKGGITVSGGEAMLQIDFITALFIEAKKLGIIHTLDTGCFYR 144

30

Query: 121 TPEYHAILEKLLDVTDLVLLDLKEIDSEQHKIVTRQSNKNILQFARYLSDRGTPVWIRHV 180
 TPEYH +L+ LL VTDL+LLDLKEID +QHKIVTRQ NKNILQFARYLSD+ PVWIRHV
 Sbjct: 145 TPEYHQVLDNLLAVTDLILLDLKEIDKQHKIVTRQPNKNILQFARYLSDKQIPVWIRHV 204

Query: 181 LVPGLTDIDDHLRLGEFVQTLNVDKFEVLPYHTMGEFKNRELGIPIYPLAGVKPPTPER 240
 LVPGLTDIDDHL RLGEFV+TL NVDKFEVLPYHTMGEFKNRELGIPI L GVKPPT ER
 Sbjct: 205 LVPGLTDIDDHLRLGEFVKTLNVDKFEVLPYHTMGEFKNRELGIPIQLEGVKPPTKER 264

35

Query: 241 VKNAKDIMKTESYTEYLKRI 260
 V+NAK++M+TESYTEY+ RI
 Sbjct: 265 VQNAKNLMQTESYTEYMNRI 284

40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1194

A DNA sequence (GBSx1270) was identified in *S.agalactiae* <SEQ ID 3717> which encodes the amino acid sequence <SEQ ID 3718>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -7.06 Transmembrane 105 - 121 (103 - 126)
 INTEGRAL Likelihood = -5.57 Transmembrane 137 - 153 (136 - 162)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50

The protein has homology with the following sequences in the GENPEPT database.

55

>GP:AAC05772 GB:AF051356 putative hemolysin [Streptococcus mutans]
 Identities = 347/445 (77%), Positives = 406/445 (90%), Gaps = 1/445 (0%)

Query: 1 MQDPGSQSLLQLQFVILLITLFFNAFFSASEMALVSLNRSKVEQKAEEGDKRYRRLLDVLE 60
 M+DPGSQSL+LQF++LLITL NAFFSA+EMALVSLNR++VEQKAEEG+K+Y RLL VLE

-1341-

Sbjct: 1 MEDPGSQLILQFLLLILTLCAFFSATEMALVSLNRARVEQKAEKEKKYIRLLKVL 60

Query: 61 NPNNFLSTIQVGITFISLLQGASLSASLGHVISGWLGN SATARTAGSIIALIFLTVVSIV 120
NPNNFLSTIQVGIT I+LL GASL+ SLG I+ W GNSATARTAGS+I+L FLTY+SIV

5 Sbjct: 61 NPNNFLSTIQVGITLITLLSGASLADSLGREIAVWFGNSATARTAGSLISLAFITYISIV 120

Query: 121 LGELYPKRIAMNLKDRLAIVSAPIIIFLGKIVSPFVWLLSASTNLLSRITPMTFDDADEK 180
LGELYPKRIAMNLK+ LA++SAP+IIFLGK+VSPFVWLLS STNLLSR+TPMTFDDADEK

10 Sbjct: 121 LGELYPKRIAMNLKENLAVLSAPVIIIFLGKIVSPFVWLLSVSTNLLSRITPMTFDDADEK 180

Query: 181 MTRDEIEYMLTNSEETLEAEIEIEMLQGIFSLDEMAREVMVPRTD AFMIDINND AQSNIE 240
MTRDEIEYMLTNSEETL+A+EIEMLQG+FSLDE+MAREVMVPRTD AFM+DIN+D+ I+

15 Sbjct: 181 MTRDEIEYMLTNSEETLDAEIEIEMLQGVFSLDELMAREVMVPRTD AFMVDINDSSDI IQ 240

Query: 241 GILSQNFSRVPVFDKDRVGVLTHTKRLL EAGFKTGFD TIDL RKILQEPLFVPETIFVD 300
IL++ FSR+PV+DDDKD+++G++HTK LL AGFK GFD I+LR+ILQEPLFVPETI V+

20 Sbjct: 241 TILNERFSRIPVYDDDKDKIIGIHTKNLLNAGFKEGFDHINLRILQEPLFVPETIVVN 300

Query: 301 DLLKALRNTQNQMAILLDEYGGVAGLVTLLEDLLEEIVGEIDDETDTAEQFVREIDENIYI 360
DLL AL+NTQNQMAILLDEYGGVAGLVTLLEDLLEEIVGEIDDET VREI +N YI

25 Sbjct: 301 DLLTALKNTQNQMAILLDEYGGVAGLVTLLEDLLEEIVGEIDDETDTAISVREIDANTYI 360

Query: 361 VLGTMTLNEFNDFETELESDDVDTIAGYYLTGVGSIPNQEKKVAYEVD SKDKHITLIND 420
VLGTMTLN+FN+YFET+LES D+VDTIAG+YLTGVG+IP+QEEK +EV+S KH+ LIND

30 Sbjct: 361 VLGTMTLNDNFNYFETDLESDNVDTIAGFYLTGVGTIPSQEEKEHFEVESNGKHLELIND 420

Query: 421 KVKDGRITKLKVLSDIEQ-NIEKD 444
KVKDGR+TKLK+L+S++E+ EKD

35 Sbjct: 421 KVKDGRVTKLKILVSEVEEKEDEKD 445

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3719> which encodes the amino acid sequence <SEQ ID 3720>. Analysis of this protein sequence reveals the following:

Possible site: 42

35 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.76	Transmembrane	22 - 38 (16 - 47)
INTEGRAL	Likelihood = -5.57	Transmembrane	118 - 134 (117 - 138)
INTEGRAL	Likelihood = -3.19	Transmembrane	150 - 166 (149 - 169)

40 ----- Final Results -----

bacterial membrane	---	Certainty=0.4503(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

45 The protein has homology with the following sequences in the databases:

>GP:AAC05772 GB:AF051356 putative hemolysin [Streptococcus mutans]
Identities = 343/443 (77%), Positives = 401/443 (90%)

50 Query: 14 MEDPVSQSLVIQFLLLVLTLLNAFFSASEMALVSLNRSRVEQKAADGDKKYARLLRVLE 73
MEDP SQL++QFLLL++LTL NAFFSA+EMALVSLNR+RVEQKA +G+KKY RLL+VLE

Sbjct: 1 MEDPGSQLILQFLLLILTLCAFFSATEMALVSLNRARVEQKAEKEKKYIRLLKVL 60

Query: 74 EPNHFLSTIQVGITFISLLSGASLSASLGKVISGWLGN SATARTAGTIIISLVFLTVVSIV 133
PN+FLSTIQVGIT I+LLSGASL+ SLG+ I+ W GNSATARTAG++ISL FLTY+SIV

55 Sbjct: 61 NPNNFLSTIQVGITLITLLSGASLADSLGREIAVWFGNSATARTAGSLISLAFITYISIV 120

Query: 134 LGELYPKRIAMNLKDKLAIVSAPIIIGLGRVSPFVWLLSASTNLLSRITPMTFDDADEQ 193
LGELYPKRIAMNLK+ LA++SAP+II LG++VSPFVWLLS STNLLSRITPMTFDDADE+

60 Sbjct: 121 LGELYPKRIAMNLKENLAVLSAPVIIIFLGKIVSPFVWLLSVSTNLLSRITPMTFDDADEK 180

Query: 194 MTRDEIEYMLSKSEATLDAEIEIEMLQGVFSLDEMAREVMVPRTD AFMIDINDDPLENIQ 253
MTRDEIEYML+ SE TLDA+EIEMLQGVFSLDE+MAREVMVPRTD AFM+DINDD + IQ

65 Sbjct: 181 MTRDEIEYMLTNSEETLDAEIEIEMLQGVFSLDELMAREVMVPRTD AFMVDINDSSDI IQ 240

Query: 254 EILKQSF SRIPVYDVKDKIIGLIHTKRLL ESGFRQGF DQINMRKMLQEPLFVPETIFVD 313

-1342-

IL + FSRIPVYD DKDKIIG+IHTK LL +GF++GFD IN+R++LQEPLFVPETI V+
 Sbjct: 241 TILNERFSRIPVYDDDKDKIIGIHTKLNAGFKEGFDHINLRRLQEPLFVPETIVVN 300

Query: 314 DLLRQLRNTQNQMAILLDEYGGVAGLVLTLEDLLEEIVGEIDDETDKAEQFVHEIGDNTYI 373
 5 DLL L+NTQNQMAILLDEYGGVAGLVLTLEDLLEEIVGEIDDETDK V EI DNTYI
 Sbjct: 301 DLLTALKNTQNQMAILLDEYGGVAGLVLTLEDLLEEIVGEIDDETDKTAISVREIADNTYI 360

Query: 374 VVGTMTLNEFNDFYFDTESDDVDVTIAGFYLTGIGTIPSQEKEAYEIDNKKHLVLIND 433
 10 V+GTMTLN+FN+YF+T+LESD+VDTIAGFYLTG+GTIPSQE+KE +E+++ KHL LIND
 Sbjct: 361 VLGTMTLNEFNDFYFDTESDDVDVTIAGFYLTGIGTIPSQEKEHFEVESNGKHLLELIND 420

Query: 434 KVKDGRITKLLKILSNIEQIIEE 456
 KVKDGR+TKLK+++S +E+ +E
 15 Sbjct: 421 KVKDGRVTKLKILVSEVEEKEDE 443

An alignment of the GAS and GBS proteins is shown below.

Identities = 364/444 (81%), Positives = 417/444 (92%)

Query: 1 MQDPGSQSLLQLFVILLILTLFNAFFSASEMALVSLNRSKVEQKAEEGDKRYRRLDVL 60
 20 M+DP QSLL++QF++L++LTL NAFFSASEMALVSLNRS+VEQKA +GDK+Y RLL VLE
 Sbjct: 14 MEDPVSQSLLVIQFLLLVLTLLNAFFSASEMALVSLNRSRVEQKAADGDKKYARLLRVLE 73

Query: 61 NPNNFLSTIQVGITFISLLQGASLSASLGHVISGWLGN SATARTAGSIIALIFLTYVSIV 120
 25 PN+FLSTIQVGITFISLL GASLSASLG VISGWLGN SATARTAG+II+L+FLTYVSIV
 Sbjct: 74 EPNHFLSTIQVGITFISLLSGASLSASLGKVISGWLGN SATARTAGTIISLVFLTYVSIV 133

Query: 121 LGELYPKRIAMNLKDR LAIVSAPIIIIFLGKIVSPFVWLLSASTNLLSRITPMTFDDADEK 180
 30 LGELYPKRIAMNLKD+LAIVSAPIII LG++VSPFVWLLSASTNLLSR+TPMTFDDADE+
 Sbjct: 134 LGELYPKRIAMNLKDKLAIVSAPIIIIGLGRLVSPFVWLLSASTNLLSRITPMTFDDADEQ 193

Query: 181 MTRDEIEYMLTNSEETLEAEIEMLQGIFSLDEM MAREVMVPRTD AFMIDINND AQSNIE 240
 35 MTRDEIEYML+ SE TL+AEIEMLQG+FSLDEM MAREVMVPRTD AFMIDIN+D NI+
 Sbjct: 194 MTRDEIEYMLSKSEATLDAEIEMLQGVFSLDEM MAREVMVPRTD AFMIDINDDPLENIQ 253

Query: 241 GILSQNFSRVPVFD DDKDRVGVGLHTRKRLLEAGFKTGFD TIDL RKILQEPLFVPETIFVD 300
 40 IL Q+FSR+PV+D DKD+++G+++HTKRLL E+GF+ GFD I++RK+LQEPLFVPETIFVD
 Sbjct: 254 EILKQSFSRIPVYDVKDKIIGLIHTKRLL ESGFRQGFQINMRKMLQEPLFVPETIFVD 313

Query: 301 DLLKALRNTQNQMAILLDEYGGVAGLVLTLEDLLEEIVGEIDDETDKAEQFVREIDENTYI 360
 45 DLL+ LRNTQNQMAILLDEYGGVAGLVLTLEDLLEEIVGEIDDETD AEQFV EI +N YI
 Sbjct: 314 DLLRQLRNTQNQMAILLDEYGGVAGLVLTLEDLLEEIVGEIDDETDKAEQFVHEIGDNTYI 373

Query: 361 VLGTMTLNEFNDFYFETELESDDVDVTIAGYYLTGVGSIPNQEKEVAYEVD SKDKHITLIND 420
 50 V+GTMTLNEFNDFYF+TELESDDVDVTIAG+YLTG+G+IP+QE+K AYE+D+KDKH+ LIND
 Sbjct: 374 VVGTMTLNEFNDFYFDTESDDVDVTIAGFYLTGIGTIPSQEKEAYEIDNKKHLVLIND 433

Query: 421 KVKDGRITKLLKVLSDIEQNIEKD 444
 KVKDGRITKLLK++LS+IEQ IE+D
 Sbjct: 434 KVKDGRITKLLKILSNIEQIIEED 457

SEQ ID 3718 (GBS70d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 120 (lane 8-10; MW 65kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 120 (lane 11 & 12; MW 44kDa) and in Figure 179 (lane 5; MW 35kDa).

GBS70d-His was purified as shown in Figure 231, lane 9-10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1343-

Example 1195

A DNA sequence (GBSx1271) was identified in *S.agalactiae* <SEQ ID 3721> which encodes the amino acid sequence <SEQ ID 3722>. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1212(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB84230 GB:AL162754 hypothetical protein NMA0960 [Neisseria
meningitidis Z2491]
Identities = 80/184 (43%), Positives = 119/184 (64%), Gaps = 3/184 (1%)

Query: 1  MIKRPIHLSDFLAEVIDKEAITLDATMGNGNDTVFLAKSSK---KVYAFDIQEEAIKAKT 57
      ++K + +H L + + + LD T GNG+DI+FLA+++ KV+AFDIQ +A+ T
Sbjct: 2  LLKNILPFAHCLLRQALPEGGNALDGTAGNGHDTLFLAQTAGIRGKVWAFDIQPQALNNT 61

Query: 58  KAKLTEQGIGSNAELILDGHENLEQYVHTPLRAAIFNLGYLPSADKTVITKPHTTIKAIAKN 117
      + +L E G S N LILDGHENL+QY+ PL AAI FN G+LP DK++ T+ T+I A+
Sbjct: 62  RCRLQEAGYSNVRLLILDGHENLKQYIPKPLDAAIFNFGWLPGGDKSLTTRTETSIAALSA 121

Query: 118 VLDILEVGGRLSLMVYYGHDGKSEKDAVIAFVEQLPQNNFATMLYQPLNQVNTPPFLIM 177
      L +L+ G L ++Y GH+ GK E +A+ + + LPQ FA + Y N+ N+PP+L+
Sbjct: 122 ALSLLKENGMLIAVLYPGHENGKQEAETQWAKNLPQEQFAVLRYSFTNRKNSPPYLLA 181

Query: 178 VEKL 181
      EKL
Sbjct: 182 FEKL 185

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3723> which encodes the amino acid sequence <SEQ ID 3724>. Analysis of this protein sequence reveals the following:

```

Possible site: 55
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -1.44      Transmembrane 127 - 143 ( 123 - 143)

----- Final Results -----
      bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9101> which encodes the amino acid sequence <SEQ ID 9102>. Analysis of this protein sequence reveals the following:

```

Possible site: 46
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -1.44      Transmembrane 118 - 134 ( 114 - 134)

----- Final Results -----
      bacterial membrane --- Certainty= 0.157(Affirmative) < succ>
      bacterial outside --- Certainty= 0.000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 124/184 (67%), Positives = 156/184 (84%)

Query: 1  MIKRPIHLSDFLAEVIDKEAITLDATMGNGNDTVFLAKSSKKVYAFDIQEEAIKAKT 60
      M+KRPIHLSDFLAEV+DK ++ +DATMGNGNDT FLA+ +KKVYAFD+QE+AI KT +

```

-1344-

Sbjct: 10 MLKRPIHLSDFLAEVVDKSSVVVDATMGNGNDTAFILAQLAKKVYAFDVQEQAIRKTSER 69

Query: 61 LTEQGISNAELIILDGHELEQYVHTPLRAAIFNLGYLPSADKTVITKPHTTIKAIKNVLD 120
L + G+SNAELIL GHE ++QYV P+RAAIFNLGYLPSADK++IT P+TT++A+ +L

5 Sbjct: 70 LAQLGLSNAELILAGHEAVDQYVTEPVRAAIFNLGYLPSADKSIITLPTTLQALSLLT 129

Query: 121 ILEVGGRLSLMVYGHGDKSEKDAVIAFVEQLPQNNFATMLYQPLNQVNTPPFLIMVEK 180
+L VGGR+++MVYGHGDK EKDA++ FV+QL Q + MLYQPLNQVNTPPFLIM+EK

10 Sbjct: 130 LLMVGGRIAIMVYGHGDKSLEKDALLDFVKQLDQRKVSAMLYQPLNQVNTPPFLIMLEK 189

Query: 181 LQSY 184
L +

Sbjct: 190 LADF 193

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1196

A DNA sequence (GBSx1272) was identified in *S.agalactiae* <SEQ ID 3725> which encodes the amino acid sequence <SEQ ID 3726>. Analysis of this protein sequence reveals the following:

20 Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.1948(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:AAC00380 GB:AF008220 YtqA [Bacillus subtilis]
Identities = 161/302 (53%), Positives = 220/302 (72%), Gaps = 4/302 (1%)

Query: 2 KKRYRAINDDYYRELFGEKIFKLPIDAGFDCPNRDGTVARGGCTFCTVSGSGDAIVAPEAP 61
+KRY +N + RE FG K+FK+ +D GFDCPNRDGTVG GGCTFC+ +GSGD

35 Sbjct: 13 EKRYHTLNVHLREHFGHKVFKVALDGGFDCPNRDGTVAHGGCTFCSAAGSGDFAGNRITD 72

Query: 62 IREQFYKEIDFMHRKWPEVNKYLTVFQNFNTNTHAKLEIKERYEQAINEPGVIGINIGTR 121
+ QF+ + MH KW + KY+ YFQ FNTNTHA +E+++E++E + V+GI+I TR

40 Sbjct: 73 LITQPHDIKRMHEKWKD-GKYIAYFQAFNTNTHAPVEVLREKFESVLALDDVVGISTATR 131

Query: 122 PDCLPDETIYYLAELSERMHVLELGLQTTYEATSALINRAHSYDLYKKTVKRIRELAPK 181
PDCLPD+ + YLAEL+ER ++ +ELGLQT +E T+ LINRAH ++ Y + V ++R+

45 Sbjct: 132 PDCLPDDVDVYLAELNERTYLVWELGLQTVHERTALLINRAHDFNCYVEGVNKLKRGH-- 189

Query: 182 VEIVSHLINGLPGETHDMMVENVRCVTDNDIQGIKLHLLHMTNTRMQRDYHEGRLRL 241
+ + SH+INGLP E DMM+E + V D D+QGIK+HLLHL+ T M + Y +G+L L

50 Sbjct: 190 IRVCSHIINGLPLEDRDMMMETAK-AVADLDVQGIKIHLLHLLKGTMPVKQYKKGKLEFL 248

Query: 242 SQEDYISIIICDQLEIIPKHIVHRTGDAPRHMLIGPMWSLNKWEVLNAIDKEMEKRQSY 301
SQ+DY+ ++CDQLEIIP +++HRTGD P ++IGPMWS+NKWEVL AI+KE+E R SY

55 Sbjct: 249 SQDDYVQLVCDQLEIIPPEMIVHRTGDGPIELMIGPMWSVNKWEVLGAINKELENRGSY 308

Query: 302 QG 303
QG

Sbjct: 309 QG 310

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3727> which encodes the amino acid sequence <SEQ ID 3728>. Analysis of this protein sequence reveals the following:

60 Possible site: 57
>>> Seems to have no N-terminal signal sequence

-1345-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2023(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 260/307 (84%), Positives = 290/307 (93%), Gaps = 1/307 (0%)

10

Query: 1 MKKRYRAINDDYYRELFGEKIFKLPIDAGFDCPNRDGTVARGGCTFCTVSGSGDAIVAPEA 60

MKKRY+ +N++YR+LFG K+FK+PIDAGFDCPNRDGTVA GGCTFCTVSGSGDAIVAP+A

Sbjct: 7 MKKRYQTLNEHYRQLFGAKMFKVPIDAGFDCPNRDGTVAHGGCTFCTVSGSGDAIVAPDA 66

15

Query: 61 PIREQFYKEIDFMHRKWPEVNKYLVIYFQNFNTNTHAKLEIIKERYEQAINEPGVGINIGT 120

PI+EQFYKEIDFMHRKWP+VN+YLVYFQNFNTNTH +++I++RYEQAINEPGV+GINIGT

Sbjct: 67 PIKEQFYKEIDFMHRKWPDVNRVYLVYFQNFNTNTHDVTVDVIRDRYEQAINEPGVGINIGT 126

20

Query: 121 RPDCLPDETIYYLAELSERMHVTLELGLQTTYEATSALINRAHSYDLYKKTVKRIRRELAP 180

RPDCLPD+TI YLAELSERMHVT+ELGLQTTYE TS LINRAHSYDLYK+TV+R+R P

Sbjct: 127 RPDCLPDDTIAYLAELSERMHVTLELGLQTTYEETSRLINRAHSYDLYKETVRRRLRHY-P 185

25

Query: 241 LSQEDYISIIICDQLEIIPKHIVIHRTGDAPRHMLIGPMWSLNKWEVLNAIDKEMEKRQS 300

LSQ+DY+SIICDQLEIIPKHIVIHRTGDAPR MLIGPMWSLNKWEVLNAIDKEME+R S

Sbjct: 246 LSQXDYVSIICDQLEIIPKHIVIHRTGDAPRDLIGPMWSLNKWEVLNAIDKEMERRGS 305

30

Query: 301 YQGCKAE 307

+QGCK +

Sbjct: 306 FQGCKVD 312

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

35 Example 1197

A DNA sequence (GBSx1273) was identified in *Sagalactiae* <SEQ ID 3729> which encodes the amino acid sequence <SEQ ID 3730>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have an uncleavable N-term signal seq

40

INTEGRAL Likelihood = -9.82 Transmembrane 10 - 26 (6 - 30)

INTEGRAL Likelihood = -4.73 Transmembrane 93 - 109 (87 - 112)

INTEGRAL Likelihood = -4.57 Transmembrane 163 - 179 (161 - 181)

INTEGRAL Likelihood = -2.97 Transmembrane 189 - 205 (185 - 205)

45

INTEGRAL Likelihood = -1.97 Transmembrane 58 - 74 (58 - 74)

INTEGRAL Likelihood = -0.75 Transmembrane 130 - 146 (130 - 146)

----- Final Results -----

bacterial membrane --- Certainty=0.4927(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA79986 GB:Z21972 ORF2 [Bacillus megaterium]

Identities = 62/159 (38%), Positives = 92/159 (56%), Gaps = 3/159 (1%)

55

Query: 34 ISFDQTIQESVRGQLPNLSTRFFKLITVIGNTVSQIAIAIMSVTFY--LKKWYFQARFI 91

+ FD+ + V+G L T K T IG+T S I ++++ + F Y LK F

Sbjct: 34 LKFDQDVISLVQGWESPLLTDMKFYTYIGSTASLIILSLVILFFLYRILKHLRLVLFPT 93

60

Query: 92 AVNAISGICILSLKLIFQVRPTLTHLVFAGGYSFPGHSMGTFMIFGSIILLQYMP 151

AV + S + L +KL FQR RP L L+ GYSFPGH+M F ++G + LL ++

-1346-

Sbjct: 94 AV-MVGSPLNLMLVKLFFQRRPDLHRLIDIGGYSFPGHAMNAFSLYGILTFLLWRHIT 152

Query: 152 KSIWKLLCQGTGLGLLIFLIGLSRIYLGVHFPTDVLGFI 190
 ++L L+I IG+SRIYLGVB+P+D++AG++

5 Sbjct: 153 ARWARILLILFSMLMILSIGISRIYLGVHYPDIAGYL 191

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1851> which encodes the amino acid sequence <SEQ ID 1852>. Analysis of this protein sequence reveals the following:

Possible site: 15
 10 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -11.30 Transmembrane 154 - 170 (150 - 181)
 INTEGRAL Likelihood = -10.88 Transmembrane 65 - 81 (58 - 93)
 INTEGRAL Likelihood = -8.97 Transmembrane 10 - 26 (5 - 31)
 INTEGRAL Likelihood = -3.77 Transmembrane 86 - 102 (86 - 105)
 15 INTEGRAL Likelihood = -2.71 Transmembrane 185 - 201 (183 - 202)
 INTEGRAL Likelihood = -1.54 Transmembrane 130 - 146 (130 - 148)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 88/197 (44%), Positives = 134/197 (67%), Gaps = 1/197 (0%)
 25 Query: 1 MLSRQNSKLIQAFIAIILFFSLGLVVIKYPDTVISFDOTIQESVRGQLPNLSTRFFKLIT 60
 M ++Q LI +F A+++F +G +K++P+ + D TIQ +RG LP + T+FF+ +T
 Sbjct: 2 MTNKQTHFLIASF-ALLIFVIIGYTVKFFPERLALLDNTIQAEIRGNLPIVLTQFFRGVT 60
 30 Query: 61 VIGNTVSQIAIAIMSVTFCYLKKWYPQARFIAVNAIISGICILSLKLIQVRVPTLTHLV 120
 V GN ++Q+ + I+SV + KW +A FI N I+ I +LKL +QR RP + HLV
 Sbjct: 61 VFGNVTQVLLVIVSVLVLFPMKWKIEALFILSNGAIAAFLITTLKLFYQRPRAIEHLV 120
 35 Query: 121 FAGGYSFPGSHMGTFMIFGSIILQYMPKSIWKLLCQGTGLGLLIFLIGLSRIYLGVB 180
 +AGGYSFPGSH+MG+ +IFGS++I+ + + + + +LI LIGLSRIYLGVB
 Sbjct: 121 YAGGYSFPGSHAMGSMIFGSLIICYQLHSHKLLQFVTSMIFILILLIGLSRIYLGVB 180
 Query: 181 FPTDVLGFIAYGILN 197
 +P+D+LAGF+L +GIL+
 40 Sbjct: 181 YPSDILAGFVLGFGILH 197

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1198

45 A DNA sequence (GBSx1274) was identified in *S.galactiae* <SEQ ID 3731> which encodes the amino acid sequence <SEQ ID 3732>. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 50 INTEGRAL Likelihood = -8.44 Transmembrane 35 - 51 (33 - 59)
 INTEGRAL Likelihood = -6.53 Transmembrane 193 - 209 (179 - 211)
 INTEGRAL Likelihood = -4.46 Transmembrane 64 - 80 (60 - 82)
 INTEGRAL Likelihood = -4.09 Transmembrane 108 - 124 (103 - 128)
 INTEGRAL Likelihood = -2.71 Transmembrane 150 - 166 (148 - 166)
 55 INTEGRAL Likelihood = -0.06 Transmembrane 174 - 190 (174 - 190)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 60 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1347-

A related GBS nucleic acid sequence <SEQ ID 9977> which encodes amino acid sequence <SEQ ID 9978> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAC83944 GB:L47648 putative [Bacillus subtilis]
    Identities = 53/186 (28%), Positives = 109/186 (58%)

    Query: 33 RKMVTIAILSALSFLMMVSFFELIPGAFLKVDLSILPMLVAFILFDLKSSYGVLLLRSL 92
              +K+V +++LS+++FVLM+++FP      ++LK+DFS +P ++A +++  + V ++++
    Sbjct: 4   KKLVVVSMSSIAFVLMLLNFFPGLPDYLDKIDFSDVPAITAILIYGPLAGIAVEAIKNV 63

10  Query: 93 LKVILANRGPETFIGLPMNMVALALFLASFAIPWKNRESAKDFIKASLFGTVSLTVSMVA 152
              L+ I+      +G N +A LF+ A +K SAK + L GT ++T+ M
    Sbjct: 64 LQYIIQGSMAQVPGQVANFIAGTLFILPTAFLFKKLSAKGLAVSLLLGTAAMTILMSI 123

15  Query: 153 LNYVFAIPLYAIFANFDIRTFIGVGNVLLTMVIPFNIVEGILISTVFYLTIVYVACLPIER 212
              LNYV +P Y F + + + ++ ++PFN+++GI+I++VF L ++ P +E+
    Sbjct: 124 LNYVLILPAYTWFLHSPALSDSALKTAVVAGILPFNMIGIVITVVFSLIFIKLKPWIEQ 183

20  Query: 213 YKKTNV 218
              + ++
    Sbjct: 184 QRSAHI 189
  
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3733> which encodes the amino acid sequence <SEQ ID 3734>. Analysis of this protein sequence reveals the following:

```

25  Possible site: 26
    >>> Seems to have a cleavable N-term signal seq.
        INTEGRAL    Likelihood = -6.48    Transmembrane    82 - 98 ( 74 - 100)
        INTEGRAL    Likelihood = -3.93    Transmembrane    161 - 177 ( 152 - 178)
        INTEGRAL    Likelihood = -3.61    Transmembrane    108 - 124 ( 107 - 126)
30  INTEGRAL    Likelihood = -3.61    Transmembrane    33 - 49 ( 31 - 50)

    ----- Final Results -----
        bacterial membrane --- Certainty=0.3590(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
  
```

The protein has homology with the following sequences in the databases:

```

40  >GP:AAC83944 GB:L47648 putative [Bacillus subtilis]
    Identities = 46/182 (25%), Positives = 97/182 (53%)

    Query: 3   KTHKMIMIGILSAISFLMLVSAFIPGAFLKIEFSIIPVLFGLMIMDLKSAYLILLR 62
              K K++++ +LS+I+F+LML++F      +LKI+FS +P + ++I + + ++
    Sbjct: 2   KVKKLVVVSMSSIAFVLMLLNFFPGLPDYLDKIDFSDVPAITAILIYGPLAGIAVEAIK 61

45  Query: 63 SILKLFLNNRGVNDFIGLEPMNIIAIALFVTAFAFVWNRQKTLNQYVFASLLGTGLTTFGM 122
              ++L+ +      +G N IA LF+ A ++ + +      + LLGT +T M
    Sbjct: 62 NVLQYIIQGSMAQVPGQVANFIAGTLFILPTAFLFKKLSAKGLAVSLLLGTAAMTIL 121

50  Query: 123 VVLNYTFAIPLYAIFANIDIRAYIGVTKYMMTMVIPFNLVEGLIFAITFYFVYIASKPIL 182
              +LNY +P Y F + + + ++ ++PFN+++G++ + F ++I KP +
    Sbjct: 122 SILNYVLILPAYTWFLHSPALSDSALKTAVVAGILPFNMIGIVITVVFSLIFIKLKPWI 181

    Query: 183 ER 184
              E+
55  Sbjct: 182 EQ 183
  
```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 110/185 (59%), Positives = 144/185 (77%)

60  Query: 29 MTNTRKMVTIAILSALSFLMMVSFFPLIPGAFLKVDLSILPMLVAFILFDLKSSYGVLL 88
              M+ T KM+ I ILSA+SF+LM+VSF +IPGA FLK++FSI+P+L ++ DLKS+Y +LL
  
```

-1348-

Sbjct: 1 MSKTHKMIMIGILSAISFLMLVSAIIPGA AFLKIEFSIIPVLFGLMIMDLKSAYLILL 60

Query: 89 LRSLLKVILANRGPETFIGLEPMNMVALALFLASFAIFWKNRESAKDFIKASLFGTVSLTV 148
 LRSLLK+ L NRG FIGLEPMN++A+ALF+ +FA+ W +++ ++ ASL GT LT

5 Sbjct: 61 LRSLLKLFNNRGVNDFIGLEPMNIIAIALFVTAFALVWNRQKTL SQYVFASLLGTGLLTF 120

Query: 149 SMVALNVVFAIPLYAIFANFDIRTFIGVGNLYLLTMVIPFNIVEGILISIVFYLTYYVACL P 208
 MV LNY FAIPLYAIFAN DIR +IGV Y++TMVIPFN+VEG++ +I FY Y+A P

10 Sbjct: 121 GMVVLNNTFAIPLYAIFANIDIRAYIGVTKYMMTMVIPFNLVEGLIFAITFYFVYIASKP 180

Query: 209 ILERY 213
 ILERY

Sbjct: 181 ILERY 185

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1199

A DNA sequence (GBSx1275) was identified in *S.agalactiae* <SEQ ID 3735> which encodes the amino acid sequence <SEQ ID 3736>. Analysis of this protein sequence reveals the following:

20 Possible site: 31
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -11.04 Transmembrane 278 - 294 (270 - 298)

25 ----- Final Results -----
 bacterial membrane --- Certainty=0.5416(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 30 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3736 (GBS150) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 7; MW 29.7kDa) and in Figure 175 (lane 4 & 5; MW 30kDa).

Purified GBS150-His is shown in Figure 110A, Figure 199 (lane 5) and Figure 227 (lanes 6-7).

- 35 The purified GBS150-His fusion product was used to immunise mice (lane 1+2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 110B), FACS (Figure 110C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 1200

A DNA sequence (GBSx1276) was identified in *S.agalactiae* <SEQ ID 3737> which encodes the amino acid sequence <SEQ ID 3738>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

45 Possible site: 40
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -15.34 Transmembrane 264 - 280 (257 - 285)
 INTEGRAL Likelihood = -7.64 Transmembrane 23 - 39 (12 - 41)

----- Final Results -----

-1349-

bacterial membrane --- Certainty=0.7135(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC13546 GB:AF019629 putative fimbria-associated protein

[Actinomyces naeslundii]

Identities = 95/271 (35%), Positives = 139/271 (51%), Gaps = 16/271 (5%)

10 Query: 29 VGLLITSYFFISNWWYNIKANNQVINFDNQTQKLNTKEINRRFELAKAYNRTLDPSRLSD 88
 +GLL +YP ++W + ++ Q + + E A AYN L + +
 Sbjct: 1 MGLL--TYPTAASWVSQYNQSKVTADYSAQVDGARP-DAKTQVEQAHAYNDALSAGAVLE 57

15 Query: 89 PYTE-----KEKKGIAEYAHMLEIAE--MIGYIDIPSIKQKLPIYAGTTSSVLEKGAGH 140
 K +YA++L+ ++ + IPSI LP+Y GT L KG GH
 Sbjct: 58 ANNHVPTGAGSSKSSIQYANILKANNEGLMARLKIPSISLDLPVYHGTADDTLKGLGH 117

20 Query: 141 LEGTSLPIGGKSSHTVITAHRGLPKAKLFTDLKLLKGGKIFYIHNIKEVLAYKVDQISVV 200
 LEGTSLP+GG+ + +VIT HRGL +A +FT+LDK+K G + EVL Y+V VV
 Sbjct: 118 LEGTSLPVGGEGRSVITGHRGLAEATMFTNLDKVKTGDSLIVEVFGEVLTYRVTSTKVV 177

25 Query: 201 KPDNFSKLLVVGKGDYATLLTCTPYSINSHRLLVRGHRIKYVPPVKEKNYLMKELQTHYK 260
 +P+ L V +GKD TL+TCTP IN+HR+L+ G RI Y P K+ K +
 Sbjct: 178 EPEETEALRVEEGKDLLTLVTCTPLGINTHRILLTGERI-YPTPAKDLAAAGKRPDVPFH 236

30 Query: 261 LYFLLSILVILVALL----YLKRKPKER 287
 ++ + + LI+V L L Y + KER
 Sbjct: 237 PWWAVGLAAGLIVVGLYLWRSGYAAARAKER 267

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3739> which encodes the amino acid sequence <SEQ ID 3740>. Analysis of this protein sequence reveals the following:

Possible site: 49

35 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -14.01 Transmembrane 225 - 241 (220 - 248)

----- Final Results -----

40 bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC13546 GB:AF019629 putative fimbria-associated protein

[Actinomyces naeslundii]

45 Identities = 94/250 (37%), Positives = 133/250 (52%), Gaps = 17/250 (6%)

50 Query: 1 VECYDRQQLSTYHKQVTQKKPSEMEVWQKAKAYNARLGIQVPDAF-----SFRD 52
 V Y ++ + Y QV +P +V ++A AYN L V +A S +D
 Sbjct: 13 VSQYNQSKVTADYSAQVDGARPDAKTQV-EQAHAYNDALSAGAVLEANNHVPTGAGSSKD 71

55 Query: 53 GIHDKNYESLLQIENNDIMGYVEVPSIKVTLPIYHYTTDEVLTGAGHLFGSALPVGGDG 112
 Y ++L+ N +M +++PSI + LP+YH T D+ L KG GHL G++LPVGG+G
 Sbjct: 72 S--SLQYANILKANNEGLMARLKIPSISLDLPVYHGTADDTLKGLGHLEGTSLPVGGE 129

60 Query: 113 THTVISAHRLPSAEMFTNLLVKKGDTFYFRVLNKVLAYKVDQILTVEPDQVTSLSGVM 172
 T +VI+ HRGL A MFTNL+ VK GD+ V +VL Y+V VEP++ +L
 Sbjct: 130 TRSVITGHRGLAEATMFTNLDKVKTGDSLIVEVFGEVLTYRVTSTKVVPEETEALRVEE 189

65 Query: 173 GKDYATLVITCTPYGVNTKRLLVRGHRIAYHYKKYQAKKAMKLVDKSRMWAEEVCAAFGV 232
 GKD TLVTCTP G+NT R+L+ G RI Y K + K A G+
 Sbjct: 190 GKDLLTLVTCTPLGINTHRILLTGERI-----YPTPAKDLAAAGKRPDVPFHFWAVGL 243

Query: 233 VIAIILVEMY 242
 +I+V +Y

-1350-

Sbjct: 244 AAGLIVVGLY 253

An alignment of the GAS and GBS proteins is shown below.

Identities = 93/192 (48%), Positives = 130/192 (67%), Gaps = 2/192 (1%)

5 Query: 52 VTNFDNQTKLNTKEINRRFELAKAYNRTLDPSRLSDPYTEKEKKGIAEYAHMLEIA--E 109
 ++ + Q + E+ ++ AKAYN L + D ++ ++ Y +L+I +
 Sbjct: 10 LSTYHKQVTQKKPSEMEEVWQKAKAYNARLGIQFVPDAFSFRDGIHDKNYESLLQIENND 69

10 Query: 110 MIGYIDIPSIKQKLPIYAGTTSSVLEKGAGHLEGTSLPIGGKSSHTVITAHRGLPKAKLF 169
 ++GY+++PSIK LPIY TT VL KGAGHL G++LP+GG +HTVI+AHRLP A++F
 Sbjct: 70 IMGYVEVPSIKVTLPTFYHYTIDEVLTKGAGHLFGSALPVGGDGTHTVISAHRGLPSAEMF 129

15 Query: 170 TDLDKLKKGKIFYIHNIKEVLAYKVDQISVVKPDNFSKLLVVKGDYATLLTCTPYINS 229
 T+L+ +KKG FY + +VLAYKVDQI V+PD + L V GKDYATL+TCTPY +N+
 Sbjct: 130 TNLNLVKKGDTFYFRVLNKVLAYKVDQILTVEPDQVTSLSGVMGKDYATLVTCTPYGVNT 189

Query: 230 HRLLVRGHRIKY 241
 RLLVVRGHRI Y
 20 Sbjct: 190 KRLLVRGHRIAY 201

SEQ ID 3738 (GBS210) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 3; MW 61kDa).

GBS210d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 2-4; MW 54kDa) and in Figure 187 (lane 9; MW 54kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lane 2-4; MW 28.7kDa) and in Figure 182 (lane 13; MW 29kDa). Purified GBS210d-GST is shown in lane 4 of Figure 237.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1201

A DNA sequence (GBSx1277) was identified in *S.agalactiae* <SEQ ID 3741> which encodes the amino acid sequence <SEQ ID 3742>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

35 Possible site: 42
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -10.61 Transmembrane 20 - 36 (15 - 40)
 INTEGRAL Likelihood = -7.27 Transmembrane 259 - 275 (258 - 277)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.5246(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC13546 GB:AF019629 putative fimbria-associated protein
 [Actinomyces naeslundii]
 Identities = 76/219 (34%), Positives = 120/219 (54%), Gaps = 12/219 (5%)

50 Query: 28 LSILLYPVVSRFYTYTIESNNQTDFFERAACKLSQKEINRRMALAQAYNDSL-----N 80
 + +L YP + + + T D+ A ++ + ++ A AYND+L+ N
 Sbjct: 1 MGLLTYPFTAASWVSQYNQSKVTADYS-AQVDGARPDAKTQVEQAHAYNDALSAGAVLEAN 59

Query: 81 VHLEDPYEKKRIQKGVAEYARMLEVSEK--IGTISVPKIGQKLPIFAGSSQEVLSKGACH 138

H+ P + +YA +L+ + + + +P I LP++ G++ + L KG GH
 Sbjct: 60 NHV--PTGAGSSKDOSSLQYANILKANNEGLMARLKIPISISLDLPVYHGTADDTLKGLGH 117
 Query: 139 LEGTSLPIGGNSTHTVITAHSGIPDKELFSNLKKLKKGDKFYIQNIKETIAYQVDQIKVV 198
 LEGTSLP+GG T +VIT H G + + +F+NL K+K GD ++ E + Y+V KVV
 Sbjct: 118 LEGTSLPVGEGGTRSVITGHRGLAEATMFTNLDKVKTCGDSLIVEVFGEVLTyrVTSTKVV 177
 Query: 199 TPDNFSDDLVLVPGHDYATLLTCTPIMINTHRLLLVRGHR I 237
 P+ L V G D TL+TCTP+ INTHR+L+ G RI
 Sbjct: 178 EPETEALRVEEGKDLLTLVTCPTPLGINTHRLLLTGER I 216

A related GBS gene <SEQ ID 8749> and protein <SEQ ID 8750> were also identified. Analysis of this protein sequence reveals the following:

The protein has homology with the following sequences in the databases:

```

33.4/53.0% over 277aa
Actinomyces
naeslundii
GP|3036999| putative fimbria-associated protein Insert characterized
ORF00563(382 - 1179 of 1479)
GP|3036999|gb|AAC13546.1|AF019629(1 - 278 of 365) putative fimbria-associated protein
{Actinomyces naeslundii}
%Match = 13.4
%Identity = 33.3 %Similarity = 53.0
Matches = 90 Mismatches = 118 Conservative Sub.s = 53
180      210      240      270      300      330      360      390
VVIMKRRQSKEA*G*SLMMYKRS*SCAYDLRVFQ*KYS*IISKSHYLGDDVKTKKIIKTKKKKKSNLPFIILFLIGLSI
                                                                : :
                                                                MGL
420      450      480      510      549      579      609
LLYPVVSRFYTTIESNNQIQDFERAAKKLSQKEINRRMALAQAYNDSL-----NVHLEDPYEKKRIQKGVAEYARML
| | | : : : | : | : : : : : | : | : : | : | : : | : | : |
LTYPTAASWSQYNQSKVTADYS-AQVDGARPDAKTQVEQAHAYNDALSAGAVLEANNHV--PTGAGSSKSDSSLOQYANIL
                20          30          40          50          60          70          80
633      663      693      723      753      783      813      843
EVS--EKIGTISVPKIGQKLPIFAGSSQEVLSKGAGHLEGTSLPIGGNSTHTVITAHSGIPDKELFSNLKCLKKGDKFYI
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KANNEGMLARLKIPIISLDDLVPYHGTADDTLTKGLGHLEGTSLPVGEGGTRSVITGHRGLAEATMFTNLDKVKVTGDSLIV
          90          100          110          120          130          140          150          160
873      903      933      963      993      1023      1053      1083
QNIKETIAYOVNDQIKVVTDPNFDLLVVPGHDYATLLTCTPIMINTHRLVVRGHRIPYKGPIDIEKLIKDGHLNTIYRYLF

```

-1352-

```

: | : | : | | | : | | | | | : | | | : | | : | : :
EVFGEVLTYRVSTSTKVVEPEETEARLVEEGKDLLTLVCTPLGINTHRIILLTGERI-YPTPAKD-LAAAGKRPDVPHPFW
      170      180      190      200      210      220      230

5  1098      1179      1209      1239      1269      1299
Y-----ISLVIIAWLLWL--IKRQRQKNR-LASVRKGIES*WEENFRKTLRNRSF*IDG*M*A*YYCS*LVF**PHILLF
: | : : | | | | | | : : : | : : | :
WAVGLAAGLIVVGLYLWRSYGAAARAKERALARARAAQEFPQPTWAEQMRIWMDDAGVEPQRWFTDLFPVPPQPSMEM
      250      260      270      280      290      300      310

```

SEQ ID 8750 (GBS212) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 4; MW 36kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 2; MW 61kDa).

Purified Thio-GBS212-His is shown in Figure 244, lane 5.

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1202

A DNA sequence (GBSx1278) was identified in *S. agalactiae* <SEQ ID 3743> which encodes the amino acid sequence <SEQ ID 3744>. Analysis of this protein sequence reveals the following:

```

20  Possible site: 29
    >>> Seems to have a cleavable N-term signal seq.
        INTEGRAL    Likelihood =-10.40    Transmembrane  680 - 696 ( 674 - 699)

25  ----- Final Results -----
        bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

30  >GP:CAA57459 GB:X81869 orf2 [Lactobacillus leichmannii]
    Identities = 84/325 (25%), Positives = 122/325 (36%), Gaps = 94/325 (28%)

    Query: 397 VNVVYTLKDKD-----KTVASVSLTKTSKGTI---DLNGIKFEVSGNF 437
           VNV + +KDKD              TV+  LTK++ T+  D G + F+ +
35  Sbjct: 236 VNVPNWIKDKDTFNVVDKPDITGIDIDASTVSIIDGLTKSTDYTVNKKDNGYQVVFKTT--- 292

    Query: 438 SGKFTGLENKSYMISERVSGYGSAINLENGKVTTITNTKDSDNPTPLNPTEPKVETHGKKF 497
           S   L  KS  I+              K T+TN  D   + T   +G
40  Sbjct: 293 SAAVQALAGKSLTITY-----KATLTNNATPDKA--IGNTATLSIGNGTNI 336

    Query: 498 VKTNEQGDRD--AGAQFVVKNSAGKYLALKADQSEGQKTLAAKKIALDEAIAAYNKLSAT 555
           T   G R+  GAQFV K+S              + KTLA + L + + N+S
45  Sbjct: 337 TSTPANGPRIYTGAQFVKKDS-----QSNKTLAGAEFQLVKVDSNGNIVSYA 384

    Query: 556 DQKGEKGITAKELIKTKQADYDAAFIEARTAYEWITDKARAITYTSNDQGQFEVTGLADG 615
           Q   +              +Y W      A TYTS+ G   + GL+
50  Sbjct: 385 TQASDG-----SYTWNDSATEATYTS DANGLVALKGLSYS 420

    Query: 616 -----TYNLEETLAPAGFAKLGNIKFVNVNQSYYITGGNIDYVANSNOKDATRVENKK 668
           +Y L E  AP G+AKL  +KF + QGS+  G+ + +  N K+
55  Sbjct: 421 DKLDGSEYALLEIQAPDGYAKLDSFVKFSITQGSF---GDSNKITIDNTKEG----- 470

    Query: 669 VTIPQTGGIGTILFTIIGLSIMLGA 693
           +P TGG G +F IG+ IM+ A
    Sbjct: 471 -LLPSTGGKGIYIFLAIGIVIMIVA 494

```

No corresponding DNA sequence was identified in *S. pyogenes*.

-1353-

SEQ ID 3744 (GBS59) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 8; MW 120kDa), in Figure 11 (lane 9; MW 100kDa) and in Figure 13 (lane 6; MW 74kDa).

GBS59-His was purified as shown in Figure 193, lane 2.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1203

A DNA sequence (GBSx1279) was identified in *S.agalactiae* <SEQ ID 3745> which encodes the amino acid sequence <SEQ ID 3746>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 25
    >>> Seems to have an uncleavable N-term signal seq
        INTEGRAL    Likelihood = -3.13    Transmembrane  870 - 886 ( 864 - 887)

15 ----- Final Results -----
        bacterial membrane --- Certainty=0.2253(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

20 >GP:AAD33086 GB:AF071083 fibronectin-binding protein I [Streptococcus pyogenes]
    Identities = 58/176 (32%), Positives = 83/176 (46%), Gaps = 19/176 (10%)

    Query: 6   KFSKILTLSLFCLSQIPLNTNVLGEST---VPENGA--KGKLVVVKTTDDQNKPLSKATFV 60
              K S +L+L+ F L + + + G S      NGA +G +KK D NKPL AT
25 Sbjct: 8   KLSFLLSLTGFI LGLLLVFILSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATSS 67

    Query: 61  LKTTAHPESEKIEKVTAELTGEATFDNLI PGDYTLSEETAPEGYKKTINQ TWQVKVESNGKT 120
              L +      + ++ T+ G      NL PG YTL EETAP+GY KT++TW V V NG T
30 Sbjct: 68  LTSKDGKGTSVQTFSTNDKGI VDAQNLQPGTYTLKEETAPDGYDRTSRTWTVTVYENGYT 127

    Query: 121 TIQNSGDKNSTIGQNQEELDKQYPPPTGIYEDTKESYKLEHVKGSVEN--GKSEAKA 174
              + +      I +      +D S +LE+ K SV + GK+E +
35 Sbjct: 128 KLVENPYNGEIIISKAGS-----KDVSSSLQLENPKMSVVSKYGKTEVSS 171
    Identities = 31/92 (33%), Positives = 49/92 (52%), Gaps = 14/92 (15%)

    Query: 725 PTITIKNEKKLGEIEFIKVDKDNKLLKLGATFELQEFNEDYKLYLPIKNNNSKVVTIGEN 784
              P+IT+ N K++ ++ F K+ DN + L A FEL+ N      N+ K+ N
40 Sbjct: 501 PSITVANLKRVAQLRFKKMSTDN--VPLPEAAFELRSSN-----GNSQKLEASSN 548

    Query: 785 --GKISYKDLKDGKYQLIEAVSPEDYQKITNK 814
              G++ +KDL G Y L E +P+ YQ++T K
    Sbjct: 549 TQGEVHFKDLTSGTYDLYETKAPKGYQQVTEK 580

```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 45 SEQ ID 3746 (GBS67) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 10; MW 140kDa), in Figure 11 (lane 10; MW 150kDa) and in Figure 12 (lane 6; MW 95.3kDa).

GBS67-His was purified as shown in Figure 192, lane 10.

- 50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1354-

Example 1204

A DNA sequence (GBSx1280) was identified in *S. agalactiae* <SEQ ID 3747> which encodes the amino acid sequence <SEQ ID 3748>. This protein is predicted to be Nra. Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2020(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9979> which encodes amino acid sequence <SEQ ID 9980> was also identified.

15 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3749> which encodes the amino acid sequence <SEQ ID 3750>. Analysis of this protein sequence reveals the following:

```

    Possible site: 58
    >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.75    Transmembrane 393 - 409 ( 392 - 409)
20
    ----- Final Results -----
      bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 122/325 (37%), Positives = 186/325 (56%), Gaps = 5/325 (1%)

30 Query: 7  LIENYLEKDILNQIKLLTLCY--DYYPSTITLDKSCHQLGLSELLIRKYCHDLITLTFNSQL 64
    LIE YLE I ++ +L+ L + Y P + + + GL+ L + YC +L F L
    Sbjct: 1  LIEKYLESSIESKQCLIVLFFKTSYLP---ITEVAEKTGLTFLQLNHYCEELNAFFPGSL 57

    Query: 65 SLNIEKSTIVYQSNQVTRQAFKYYHQSHVLQLLKFLITNDSGRLPLTYFSEKFGLSCA 124
    S+ I+K I Q +E +Y S+VLQLL FLI N S PLT F+ LS +
35 Sbjct: 58 SMTIQKRMISCQFTHPFKETLYLYQYASSNVLQLLAFLIKNGSHSRPLTDFARSHFLSNS 117

    Query: 125 TAYRIRKHISPLEKLGFQIVKNTITGDEYRIRYLI AFLNAQFGIEVYPMSKMDKLLIKR 184
    +AYR+R+ + PLL ++ KN I G+EYRIRYLI L ++FGI+VY +++ DK I
40 Sbjct: 118 SAYRMREALIPLLRNFELKLSKNKIVGEEYRIRYLIALLYSKFGIKVYDLTQODKNTIHS 177

    Query: 185 LLEHSTTFTASHYFPNTFIFFDTLTSLSWKRINYNVVPYSSLFTELQNIIFYDTLQYC 244
    L ST S + +F F+D LL+LSWKR ++V +P + +F +L+ +F+YD+L+
    Sbjct: 178 FLSSHSTHLKTSFPLSESFSFYDILLALSWKRHQFSVTIPQTRIFQQLKKLFVYDSLKKS 237

45 Query: 245 VKNVIIDSFKINLKKDDIDYIFLAYLTSHNSFSNPNTWTEKRIDNVIAIFENYPKFQKLLQ 304
    ++I ++N D+DY++L Y+T++NSF++ WT + I +FE F+ LL
    Sbjct: 238 SHDIETCYQLNFSAGDLDYLYLIYITANNSFASLQWTPEHIRQVCQLFEENDTFRLLLN 297

    Query: 305 PLKDALPLSGSYHDELVKVAIFFSE 329
50 P+ LP LVK +FFS+
    Sbjct: 298 PIITLLPNLKEQKASLVKALMFFSK 322

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1355-

Example 1205

A DNA sequence (GBSx1281) was identified in *S.agalactiae* <SEQ ID 3751> which encodes the amino acid sequence <SEQ ID 3752>. This protein is predicted to be galactosyltransferase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 21
    >>> Seems to have no N-terminal signal sequence (or aa 1-22)

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.1168(Affirmative) < succ>
10   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAB99071 GB:U67549 galactosyltransferase isolog [Methanococcus
    jannaschii]
    Identities = 108/395 (27%), Positives = 196/395 (49%), Gaps = 28/395 (7%)

    Query: 4   KVKTAVFSGYYLPFLGGIERYTDKMTADLVK-RGYRVVIVTTNHGDLPIIDEDKGR--- 59
              K+K + +F GYY+P +GG+E + D+ T L + Y + I N +P E + R
20   Sbjct: 3   KIKLI-IFPGYYIPHIGGLETHVDEFTKHLSEDENYDIYIFAPN---IPKYKEFEIRHNN 58

    Query: 60   -KIYRLPTKNIVKQRYPIINK-NREYNTLMKYVSDENIDFVICNTRFQLTTLLEGLSFAKN 117
              K+YR P I+ YP+ N N ++ + + + D V+ TRF TL G FAK
25   Sbjct: 59   VKVYRYPAFEIIPN-YPVPNIFNFKFWRMFFNLYKIDFDIVMTRTRFFSNTLLGFIFAKL 117

    Query: 118   HHLPS--IVLDHGSSSHFSVNNRFLDFFGAIYEHLLTARVKHYRPDFYAVSKRSVEWLKHF 175
              I ++HGS+ + + F + Y+ + + A+SK ++
    Sbjct: 118   RFKKKKLIHVEHGSAFVKLESEFKNKLSYFYDKTIGKLIFFKADYVVAISKAVKNFILEN 177

30   Query: 176   NIEAKGV--IYNSVS---ESLGSDFAGTAYLEKSADDIFTYAGRIIEKEKGIELLLEAF 229
              + K + IY + ES+G D EK + I + + GR+ K KG+E +++A+
    Sbjct: 178   FVNDKIDPIIYRGLEIEKIESIGED---KKIKEKFKNKIKLCFVGRLYKWKGVENI IKAY 234

    Query: 230   S--MSQYSENVYLQIAGDGPPELAHLKE---KYQSKQINFLGKLNFEQIMSLMAQTDIFVY 284
              E + L + G G +L LK+ Y + I F GK++FE+ +++ +DI+++
35   Sbjct: 235   VDLPKDLKEKIIILIVVGYGEDLERLKLKLAGNYLNNGIYFTGKVDFEKAIAIVKASDIYIH 294

    Query: 285   PSMYPEGLPTSILEAGLLSSAIIATDRGGTVEVIDSPELGIIMEENT-QSLHESLDLLVK 343
              S GL +S+L+A AI+A+ G EV+ GI++++N+ + + + L++
40   Sbjct: 295   SSYKGGGLSSSLQAMCCGKAIVASPYEGADEVVIDGYNGILLKDNSPEBIKRGIIKLE 354

    Query: 344   DKALREKLQONIAKRIKEHFTWEKTVEKLDYIIQK 378
              + LR+ +N IKE+F W+K+V++ I ++
45   Sbjct: 355   NNNLRKIYGENAKNFIKENFNWKKSVKEYKKIFER 389

```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3752 (GBS258) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 2; MW 43kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 7; MW 67.9kDa).

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1206

A DNA sequence (GBSx1282) was identified in *S.agalactiae* <SEQ ID 3753> which encodes the amino acid sequence <SEQ ID 3754>. Analysis of this protein sequence reveals the following:

55 Possible site: 31

-1356-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1182(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB52237 GB:Z98171 EpsQ protein [Streptococcus thermophilus]
 Identities = 112/278 (40%), Positives = 163/278 (58%), Gaps = 2/278 (0%)

Query: 1 MKYLAGIVTFNPNIERLDQNIRAIYPQVSHIYIVDNGSKNKEEISQLVADYNEEGHLTVTD 60
 M AGIV FNP+I+RL +NI A+ Q +H+Y+VDNGS N +E+ L+ YN+ +++
 15 Sbjct: 1 MDISAGIVLFNPDIKRLKENIDAVIIQCTHLYLVLDNGSGNVDEVKGLLNQYNQS-KISIL 59

Query: 61 YLTENKGIAYALNCIGQFAVAQEFDFWFLTLTDQDSVVLGDLIDNYENYLHLPKVGMLSCLY 120
 + EN+GIA ALN + A + FDW LTLDQDSVV +++ +E Y++' VG+L +
 15 Sbjct: 60 WNRENQGIKALNQLTSAAQKEGFDWILTLTDQDSVVPENIVGEFEKYINNSSVGILCPIL 119

20 Query: 121 QDMNRENLMQEFDYKEIEECITSAALMKTSTVFETSGFABEMFIDFVDSEMNRYRLSEM 180
 D N++ + D EI+ECITS +L+ + E GF E MFID VD ++ YRL + G
 Sbjct: 120 CDRNKDEEIKINEDCTEIDECITSGSLLNIKAWSEIGGFDERMFIDGVDFDICRYLRQRG 179

25 Query: 181 YKTYQVNFIFGLLHEIGHSSRVKKFGHVPHVLNHSPPFRKYYMIRNAIYIICKYGGKKRYKY 240
 YK Y ++ + LLHE+GH + V NHS FRKYY+ RN IY KK
 Sbjct: 180 YKIYCIHSVVLHELGHIEYHRFLFWKVLVKNHSAFRKYYIARNIYIYAKRRSTLLVVK 239

30 Query: 241 LVFMRNEFVRVLV-AEEQSKKIVAMIKGLKDGLLMKV 277
 + + + +++ EE K KI + +G+ DG KV
 Sbjct: 240 GLLQEIKLIGIVIFYEEDKLNKIRCTCRGIYDGFKGKV 277

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1207

A DNA sequence (GBSx1283) was identified in *S.agalactiae* <SEQ ID 3755> which encodes the amino acid sequence <SEQ ID 3756>. This protein is predicted to be EpsU protein (rfbX). Analysis of this protein sequence reveals the following:

40 Possible site: 54
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.44	Transmembrane	357 - 373 (352 - 387)
INTEGRAL	Likelihood = -7.59	Transmembrane	88 - 104 (79 - 107)
INTEGRAL	Likelihood = -7.32	Transmembrane	440 - 456 (433 - 465)
INTEGRAL	Likelihood = -6.48	Transmembrane	246 - 262 (245 - 263)
45 INTEGRAL	Likelihood = -4.78	Transmembrane	294 - 310 (290 - 312)
INTEGRAL	Likelihood = -3.88	Transmembrane	164 - 180 (162 - 183)
INTEGRAL	Likelihood = -3.56	Transmembrane	144 - 160 (136 - 161)
INTEGRAL	Likelihood = -2.87	Transmembrane	317 - 333 (316 - 334)
INTEGRAL	Likelihood = -2.71	Transmembrane	374 - 390 (374 - 393)
50 INTEGRAL	Likelihood = -0.96	Transmembrane	44 - 60 (44 - 62)
INTEGRAL	Likelihood = -0.80	Transmembrane	15 - 31 (15 - 32)

----- Final Results -----

55 bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB52225 GB:Z98171 EpsU protein [Streptococcus thermophilus]

-1357-

Identities = 189/462 (40%), Positives = 313/462 (66%)

Query: 1 MKLLKNMFYNTSYQLLTLLPLVTPVVSRLSPQIGIGINAYTSSIVMYFTLFGALGISL 60
 M+++KN YN YQ+ +++PL+T+PY+SR+L P GIGIN+YT+SIV YF LFG++G+ L
 5 Sbjet: 1 MQIVKNLYNAIYQVFIIIVPLLTIPYLSRILGPSGIGINSYTNISIVQYFVLFGSIGLGL 60

Query: 61 YGNREIAFVQSNKYKRSKIFWELVVLKLASVSIATLLFFGFVLLTNEWQLFYLIQGINLL 120
 YGNR+IAFV+ N+ K SK+F+E+ +L+L ++ +A LF F+++ ++ +YL Q I ++
 10 Sbjet: 61 YGNRQIAFVRDNQVKMSKVFEIIFILRLFTICLAYFLFVAFLIINGQYYAYLSQSIATV 120

Query: 121 ATATDISWYFIGVEDFKIIVIRNTIVKLITVVLTFVLVVKTPDDDLALYMFLLAFASLLGNL 180
 A A DISW F+G+E+FK+IV+RN IVKL+ + FL VK+ +DL +Y+ + ++L+GNL
 15 Sbjet: 121 AAAPDISWAFMGIEFNKVIIVLRNFIVKLLALFSIFLKVSYNDLNIYILITVLTSLIGNL 180

Query: 181 TVWHHLKHEIHKIPFSRLDILHLRPTLMFLPQITMQIYLSLNKSMGLGAMDSVVSAGYF 240
 T + L ++K+ + L + HL+ +L++F+PQI +QIY LNK+MLG++DSV S+G+F
 20 Sbjet: 181 TFFPSLHRYLVKVNRELRPIKHLKQSLVMFIPQIALQIYVVLNKTMLGSLDSVTSSGFF 240

Query: 241 DQSDKIIRILFTIVSAIGGVFLPRLSSLFSSGKEKQAKALLLKLVDLSNAISMLMIAGVV 300
 DQSDKI++++ IV+A G V LPR+++ F+ + + K + +AIS+ M+ G++
 25 Sbjet: 241 DQSDKIVKLVLAIVTATGTVMPLPRVANAFHREYSKIKEYMYAGFSFVSAISIPMMFGLI 300

Query: 301 GVSSTFAVFFFGKGYEAVGPLMAVESLMIICISYGNALGTQYLLASRRTKAYTMSAVIGL 360
 ++ F FF + V P++ +ES+ II I++ NA+G QYLL + + K+YT+S +IG
 30 Sbjet: 301 AITPKFVPLFFTSQFSDVIPVLMIESIAIFIWNAIGNQYLLPTNQNKSYTVSVIIGA 360

Query: 361 VANVVLNILLIPILGAMGAIISTVITEFIVSLYQAISLRDVFTFKELTRGMLRYLIAATL 420
 + N++LNI LI LGA+GA I+TVI+E V++YQ + L + +YLIA +
 35 Sbjet: 361 IVNLMLNIPLIYLGAVGASIAITVISEMSVTVYQLFIIHKQLNLHTLFSDDLKYLIAGLV 420

Query: 421 SGAVLYYINTQMSVSLVNYVIQSLVAVTIYVGIVFITKAPVI 462
 +++ I+ S + +++ V + IY+ ++ KA +I
 40 Sbjet: 421 MFLIVFKISLLTPTSWIFILLEITVGIIYIVLLIFLKAETI 462

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1208

40 A DNA sequence (GBSx1284) was identified in *S.agalactiae* <SEQ ID 3757> which encodes the amino acid sequence <SEQ ID 3758>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1742(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1209

55 A DNA sequence (GBSx1285) was identified in *S.agalactiae* <SEQ ID 3759> which encodes the amino acid sequence <SEQ ID 3760>. Analysis of this protein sequence reveals the following:

-1358-

Possible site: 25
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

10 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1210

A DNA sequence (GBSx1286) was identified in *S.agalactiae* <SEQ ID 3761> which encodes the amino acid sequence <SEQ ID 3762>. Analysis of this protein sequence reveals the following:

Possible site: 34
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.56	Transmembrane	214 - 230 (210 - 236)
INTEGRAL	Likelihood = -10.03	Transmembrane	364 - 380 (361 - 386)
20 INTEGRAL	Likelihood = -7.96	Transmembrane	272 - 288 (271 - 291)
INTEGRAL	Likelihood = -6.95	Transmembrane	23 - 39 (20 - 41)
INTEGRAL	Likelihood = -5.57	Transmembrane	191 - 207 (189 - 209)
INTEGRAL	Likelihood = -5.15	Transmembrane	434 - 450 (425 - 451)
INTEGRAL	Likelihood = -4.25	Transmembrane	143 - 159 (138 - 162)
25 INTEGRAL	Likelihood = -3.13	Transmembrane	167 - 183 (166 - 186)
INTEGRAL	Likelihood = -1.44	Transmembrane	400 - 416 (400 - 416)
INTEGRAL	Likelihood = -1.33	Transmembrane	333 - 349 (333 - 349)
INTEGRAL	Likelihood = -0.80	Transmembrane	232 - 248 (232 - 251)

30 bacterial membrane --- Certainty=0.5225(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1211

40 A DNA sequence (GBSx1287) was identified in *S.agalactiae* <SEQ ID 3763> which encodes the amino acid sequence <SEQ ID 3764>. This protein is predicted to be rhamnosyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.1792(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

-1359-

A related GBS nucleic acid sequence <SEQ ID 9981> which encodes amino acid sequence <SEQ ID 9982> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAF18951 GB:AF155805 Cps9H [Streptococcus suis]
    Identities = 53/116 (45%), Positives = 75/116 (63%), Gaps = 4/116 (3%)

    Query: 6  VLMATYNGQGFIHQDLSIRNQTLRPDYVLMRDDGSTDDTVKVVEDYIKEHRLDGWSITS 65
              VLMATYNG FI QDLSIRNQ++ D V++ DD STDDT+K+++DYIK++ LD W ++
10  Sbjct: 4  VLMATYNGSPFIKQLDSIRNQSVSADKVIWDDCSTDDTIKIIKDYIKKYSLSWVVSQ 63

    Query: 66  NDKNLGWRLNFRQLLIDVLAYEVDYVFFSDQDDTWYHHKMKMQVDIMEERQDINLL 121
              N N G F L + VFFSDQDD W HK + + I +R+++++
10  Sbjct: 64  NKSNGGHYQTFINL---TKLVQEGIVFFSDQDDIWDCHKIETMLPIF-DRENVSMV 115

```

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1212

A DNA sequence (GBSx1288) was identified in *S.galactiae* <SEQ ID 3765> which encodes the amino acid sequence <SEQ ID 3766>. This protein is predicted to be rhamnosyltransferase. Analysis of this protein sequence reveals the following:

```

20  Possible site: 13
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
25  bacterial cytoplasm --- Certainty=0.1278(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

30 A related GBS nucleic acid sequence <SEQ ID 9983> which encodes amino acid sequence <SEQ ID 9984> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

35  >GP:AAF18951 GB:AF155805 Cps9H [Streptococcus suis]
    Identities = 57/146 (39%), Positives = 81/146 (55%), Gaps = 8/146 (5%)

    Query: 10  VLMATYNGEIFISEQLDSIRQQLKPDYVLLRDCSTDETNNVNNYIAKHELEGWKIVK 69
              VLMATYNG FI +QLDSIR Q++ D V++ DDCSTD+T+ ++ +YI K+ L+ W + +
40  Sbjct: 4  VLMATYNGSPFIKQLDSIRNQSVSADKVIWDDCSTDDTIKIIKDYIKKYSLSWVVSQ 63

    Query: 70  NDKNLGWRLNFRQLLIDVLAYEVDYVFFSDQDDIWDKNERQFAIMSDKPQIEVLSADV 129
              N N G F L + VFFSDQDDIW K E I D+ + + V
40  Sbjct: 64  NKSNGGHYQTFINL---TKLVQEGIVFFSDQDDIWDCHKIETMLPIF-DRENVSM---V 115

    Query: 130  DIKTMSTEASVPFLTFSSSDRISQY 155
              K+ + + + +SDRI+ Y
45  Sbjct: 116  FCKSRLIDENGNISSPDTSRINTY 141

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1213

50 A DNA sequence (GBSx1289) was identified in *S.galactiae* <SEQ ID 3767> which encodes the amino acid sequence <SEQ ID 3768>. This protein is predicted to be dTDP-glucose 4-6-dehydratase (galE). Analysis of this protein sequence reveals the following:

-1360-

Possible site: 44

>>> Seems to have an N-terminal signal sequence

INTEGRAL Likelihood = -2.02 Transmembrane 250 - 266 (250 - 266)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.1808(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 A related GBS nucleic acid sequence <SEQ ID 9985> which encodes amino acid sequence <SEQ ID 9986> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAC14890 GB:AJ295156 d-TDP-glucose dehydratase [Phragmites
australis]

15 Identities = 108/327 (33%), Positives = 170/327 (51%), Gaps = 22/327 (6%)

Query: 29 ANKGVLSIGSNSMLASYMVFLAYLNETRNYQTQIIATARNIEKARDKFSDDLGVGKDYFTL 88

AN +L++G + S++V L N + ++I ++D +G F L

20 Sbjct: 33 ANLRLLVTGGAGFIGSHLVDKLM-----ENEKHEVIVADNFFTGSKDNLKKWIGHPRFEL 87

Query: 89 IPYDVEERLEYDGKVDYIIHAASNASPTAILSNPVSIIKANTIGTLNLLDFAKEKTIENF 148

I +DV + L + VD I H A ASP NPV IK N IGTLN+L AK +

25 Sbjct: 88 IRHDTIQPLLVE--VDQIYHLACPASPIFYKHNPVKTIKTNVIGTLNMLGLAK-RVGARI 144

Query: 149 LFLSTREVGTSIKEVIDEEAYGGFDILATRACYPESKRMAETLLQSYDQYKVPFTIAR 208

L ST EVYG ++ E +G + + R+CY E KR+AETL+ Y+ Q+ + IAR

30 Sbjct: 145 LLTSTSEVYGDPLEHPQTEAYWGNVNPIGVRSCYDEGKRVAETLMPDYHRQHGIETRIAR 204

Query: 209 IAHSFGPGMELGNDGRIMNDLLSNVIDGKDIVLKSSGTAERAFCYLADAVSGLFTILLNG 268

I +++GP M + +DGR++++ ++ + G + ++ GT R+FCY+AD V GL L+NG

35 Sbjct: 205 IFNTYGPFRMNI-DDGRVVSNFIAQAVRGDELTVQKPGTQTRSFQYVADMVDGLIK-LMNG 262

Query: 269 EVGQAYNVANEDQPIMIKDLAKLVDLFSKDNISVVFDPKTMASAGYSKMGRTTR--LTM 325

N+ N + M+ +LA+K+ +L + ++ TM+ R R +T

40 Sbjct: 263 NNTGPINLGNPGEFTML-ELAEKVKELINP-----EVTVMTEPTDDPRQRKPDITK 314

Query: 326 AKLEALGWKREVSLESIGLKTQAFEE 352

AK E LGW+ +V L G++ F E

45 Sbjct: 315 AK-EVLGWEPKVLRDGLVLMEDDFRE 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1214

45 A DNA sequence (GBSx1290) was identified in *S.agalactiae* <SEQ ID 3769> which encodes the amino acid sequence <SEQ ID 3770>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have an uncleavable N-term signal seq

50 ----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 9987> which encodes amino acid sequence <SEQ ID 9988> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11866 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]

-1361-

Identities = 77/231 (33%), Positives = 131/231 (56%), Gaps = 6/231 (2%)

Query: 13 VIFAGGVGRRMNTKGKPKQFLEVHGKPIIVHTIDIFQNTAIDAVVVVVCVSDWLDYMN 72
 VI A G G+RM G+ K F+E+ G P+I+HT+ +F + D +++V ++ L
 5 Sbjct: 6 VIPAAGQGKRMKA-GRNKLFIELKGDPIIHTLRVFDShrQCDKIILVINEQEREHFQQL 64

Query: 73 VERFNLTKVKAVVAGGETGQMSIFKGLEAAEQLATDDAVVLIHDGVRPLINEEVINANIQ 132
 + + +VAGG+ Q S++KGL+A +Q + +VL+HDG RP I E I+ I
 10 Sbjct: 65 LSDYPPQTSIELVAGGDERQHSVYKGLKAVKQ---EKIVLVHDGARPFKHEQIDELIA 120

Query: 133 SVKETGSAVTSVRAKETVVLVNDSSKISEVVDRTSRFIAKAPQSFYLSDILSVERDAISK 192
 ++TG+A+ +V K+T+ V D ++SE ++R+ + + PQ+F LS ++ +A K
 15 Sbjct: 121 EAEQTGAAILAVPVKDTIKRVQDL-QVSETIERSSSLWAVQTPQAFRLSLMKAAHAEAEK 179

Query: 193 GITDAIDSTLMGMYNRELITVEGPIYENIKITTPDDFYMFKALYDARENEQ 243
 G D+S + M + +VEG Y NIK+TTPDD +A+ ++ +
 Sbjct: 180 GFLGTDDASLVEQMEGGSVRVVEGSYTNIKLTTPDDLTSAEAIMESSEGNK 230

No corresponding DNA sequence was identified in *S.pyogenes*.

20 SEQ ID 3770 (GBS647) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 9 & 10; MW 55.9kDa + lane 8; MW 27kDa) and in Figure 186 (lane 5; MW 56kDa).. It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 12; MW 31kDa), in in Figure 140 (lane 9; MW 31kDa) and in Figure 178 (lane 6; MW 31kDa).

25 Purified GBS647-GST is shown in Figure 243, lane 4; purified GBS647-His is shown in Fig.229, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1215

30 A DNA sequence (GBSx1291) was identified in *S.agalactiae* <SEQ ID 3771> which encodes the amino acid sequence <SEQ ID 3772>. This protein is predicted to be LicD1. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2647(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9989> which encodes amino acid sequence <SEQ ID 9990> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD37094 GB:AF106539 LicD2 [Streptococcus pneumoniae]
 Identities = 85/271 (31%), Positives = 130/271 (47%), Gaps = 15/271 (5%)
 45 Query: 1 MKEMTVSEIREVQLEMLAYIDKVARDNKIEYSLGGGSLGAMRHKGFI PWDDDDIDMLER 60
 M+ + EI+E+QL +L YID+ + + I Y L G++LGA+RHKG IPWDDDDID+ L R
 Sbjct: 1 MQYLEKKEIKEIQALLDYIDETCKKHDI PYFLSYGTMLGAIRHKGMIPWDDDDIDISLYR 60

50 Query: 61 SQYERLMKALADANNSDFKLLHHSVEKNLW---PFAKLYHTKSMYLSKTDRIHPWTGIFI 117
 YERL+K + + N+ +K+L S + + W FA + T ++ T +FI
 Sbjct: 61 EDYERLLKIIIEENHPRYKVL--SYDTSSWYFHNFA SILDSTVIEDHVYKVRHDTSLFI 118

-1362-

Query: 118 DIFPLDRLPESAERQRFKKVHSAANLMCTTYPNFASGSRKLYANARLILGLP-RFIA 176
 D+FP+DR + + + + + A L G KL RL RF+
 Sbjct: 119 DVFPIDRFTDLSIVDKSY---KYVALRQLAYIKKSRVHGDGSKLKDFLRLCSWYALRFVN 175

5 Query: 177 YHGQAKKRAEIVDQVMETYNQEVPMGYTD-SRYRLKEYFPREIFSEYEDVMFENIKTR 235
 KK +DQ+++ Y G + +KE FP + F E FE
 Sbjct: 176 PRYFYKK---IDQLVKNVINTPQYEGGVGIGKEGMKEIFPVDTFKELILTEFEGRMPL 231

10 Query: 236 KIKNEHAYLNQLYGGSYMEPLPPESKRESHSY 266
 K +L Q+Y G YM P + +E +S+
 Sbjct: 232 VPKKYDQFLTQMY-GDYMTTPSKEMQEWYSH 261

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 15 vaccines or diagnostics.

Example 1216

A DNA sequence (GBSx1292) was identified in *S.agalactiae* <SEQ ID 3773> which encodes the amino acid sequence <SEQ ID 3774>. Analysis of this protein sequence reveals the following:

20 Possible site: 18
 >>> May be a lipoprotein
 INTEGRAL Likelihood = -12.05 Transmembrane 554 - 570 (547 - 575)
 ----- Final Results -----
 25 bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 SEQ ID 3774 (GBS182d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 184 (lane 8; MW 62kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1217

35 A DNA sequence (GBSx1293) was identified in *S.agalactiae* <SEQ ID 3775> which encodes the amino acid sequence <SEQ ID 3776>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4653(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1363-

Example 1218

A DNA sequence (GBSx1294) was identified in *Sagalactiae* <SEQ ID 3777> which encodes the amino acid sequence <SEQ ID 3778>. This protein is predicted to be DOLICHYL-PHOSPHATE MANNOSE SYNTHASE RELATED PROTEIN. Analysis of this protein sequence reveals the following:

5 Possible site: 29
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -2.92 Transmembrane 232 - 248 (231 - 248)

 ----- Final Results -----
 10 bacterial membrane --- Certainty=0.2168(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 A related GBS nucleic acid sequence <SEQ ID 9991> which encodes amino acid sequence <SEQ ID 9992> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC35924 GB:AF071085 putative glycosyl transferase [Enterococcus
 faecalis]
 Identities = 118/240 (49%), Positives = 152/240 (63%), Gaps = 1/240 (0%)
 20 Query: 14 KILLVIPAYNEEGSIKTVQTIIVDFKASRS-LPFELDYIVINDGSTDGTPELLDRLGLNH 72
 K+LL+IPAYNEE +I +T+ +I FK + ELDY+VINDGSTDGT ++L+ +N
 Sbjct: 2 KVLLIIPAYNEENILRTIASIETFKQEVTHFQHELDYVIVINDGSTDGTQILEVNVQINA 61

 25 Query: 73 IDLVQNLGIGGCVQTGYLYANRNHYDVAVQFDGQGQHDIRSIEDVVMPIINDEADVFVIGS 132
 I LV NLGIGG VQTGY YA N YDVA QFDGDG HDI S+ ++ P+ F GS
 Sbjct: 62 IHLVLNLGIGGAVQTGYKYALENEYDVAXQFDGDGXHDIXSLPILLEPLAEGKXCFXSXGS 121

 30 Query: 133 RFVDKKHQNFQSTAMRRRLGINLISAAIKLTGCHKVYDTSQGYRAANAALIAVLSCHYPVQ 192
 RF+ +FQS MRR GI L+S G +Y T G RA N +IA+ + YP
 Sbjct: 122 RFIPGNXASFQSKMRRXGIRLLSFCXXXAXGXTIYKVTXGKRAGNRKVIAFFAKRYPTN 181

 35 Query: 193 YPEPESTARILKKGYRLKEVTANMFEREAGTSSISSLSKIFVMTDVLTSIIAGFIKEDD 252
 YPEPES ++KK + + E NM ER G SSI +L S+ YM +V ++I+IA F+KE D
 Sbjct: 182 YPEPESIVHLIKKRFVIVERPVNMERLGGVSSIRALASVKYMLEVGSAILIAPFMKEGD 241

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3779> which encodes the amino acid sequence <SEQ ID 3780>. Analysis of this protein sequence reveals the following:

40 Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.80 Transmembrane 211 - 227 (211 - 227)

 ----- Final Results -----
 45 bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC35924 GB:AF071085 putative glycosyl transferase [Enterococcus
 faecalis]
 Identities = 104/233 (44%), Positives = 134/233 (56%), Gaps = 9/233 (3%)
 50 Query: 1 VKLLIIPAYNESSNIVNTIRTIESDAPD-----FDYIIIDDCSTDNTLAICQKQGFN 53
 +K L+IIPAYNE NI+ TI +IE+ + DY++I+D STD T I + N
 55 Sbjct: 1 MKVLLIIPAYNEENILRTIASIETFKQEVTHFQHELDYVIVINDGSTDGTQILEVNVQIN 60

 Query: 54 VISLPINLGIGGAVQTGYRYAQRQGYDVAVQVDGQGNPCYLEKMVEVLVQSSVNMVIG 113
 I L +NLGIGGAVQTGY+YA YDVA Q DGDG H+ L ++E L + G
 Sbjct: 61 AIHLVLNLGIGGAVQTGYKYALENEYDVAXQFDGDGXHDIXSLPILLEPLAEGKXCFXSXG 120

-1364-

Query: 114 SRFI--TKEGFQSSFARRIGIKYFTWLIALLTGKKITDATSGLRLIDRSLIERFANHYPD 171
 SRFI FQS RR GI+ ++ G I T G R +R +I FA YP
 Sbjct: 121 SRFIPGNXASFQSKMRRXGIRLLSFCXXXAGXTTYXVTVGXGRAGNRKVIAFFAKRYPT 180

Query: 172 DYPEPETVVDVLVSHFKVKEIPVVMNERQGGVSSISLTKSVYYMIKVTILAILV 224
 +YPEPE++V ++ F + E PV M ER GGVSSI SV YM++V AIL+
 Sbjct: 181 NYPEPESIVHLIKRKFVIVERPVNMERLGGVSSIRALASVKYMLEVGSAILI 233

10 An alignment of the GAS and GBS proteins is shown below.

Identities = 105/231 (45%), Positives = 142/231 (61%), Gaps = 8/231 (3%)

Query: 14 KILLVIPAYNEEGSIAKTVQITVDFKASRLPFELDYIVINDGSTDGTPELLDRLGLNHI 73
 K L++IPAYNE +I T++TI S + DYI+I+D STD T + + G N I
 Sbjct: 2 KKLIIIPAYNESSNIVNTIRTI-----ESDAPDFDYIIIDDCSTDNTLAICQKQGFNVI 55

Query: 74 DLVQNLGIGGCVQTYLYANRNHYDVAVQFDGQGHDIRSIEDVVMPIILDEADFVIGSR 133
 L NLGIGG VQTY YA R YDVAVQ DGDGQH+ +E +V ++ + VIGSR
 Sbjct: 56 SLPINLGIGGAVQTYRYAQRQGYDVAVQVDGQGHNPCCYLEKMVEVLVQSSVMVIGSR 115

Query: 134 FVDKHKQNFQSTAMRRGLINLISAAIKLTGCHKVYDTSGYRAANAALIAVLSCHYPVQY 193
 F+ K + FQS+ RR+GI + I L TG K+ D TSG R + +LI + HYP Y
 Sbjct: 116 FITK--EGFQSSFARRIGIKYFTWLIALLTGKKITDATSGLRLIDRSLIERFANHYPDDY 173

Query: 194 PEPESTARILKKGRLKEVTANMFEREAGTSSISLKSIFYMTDVLTSIII 244
 PEPE+ +L +++KE+ M ER+ G SSIS KS++YM V +I++
 Sbjct: 174 PEPETVVDVLVSHFKVKEIPVVMNERQGGVSSISLTKSVYYMIKVTILAILV 224

30 A related GBS gene <SEQ ID 8751> and protein <SEQ ID 8752> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
 McG: Discrim Score: 0.29
 GvH: Signal Score (-7.5): -4.34
 Possible site: 29

>>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 1 value: -2.92 threshold: 0.0
 INTEGRAL Likelihood = -2.92 Transmembrane 222- 238(221 - 238)
 PERIPHERAL Likelihood = 4.40 4
 modified ALOM score: 1.08

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.2168(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF00548(340 - 1056 of 1359)
 GP|3608398|gb|AAC35924.1||AF071085(2 - 241 of 241) putative glycosyl transferase
 {Enterococcus faecalis}
 %Match = 24.7
 %Identity = 49.2 %Similarity = 64.2
 Matches = 118 Mismatches = 85 Conservative Sub.s = 36

249 279 309 339 369 399 429 456
 L*QD*GGYGNMVIKINLSIKLCLNG*XQQIIXIRDKMMKILLVIPAYNEEGSIAKTVQITVDFKASRS-LPFELDYIV
 :||:||||||| :| :| :| || : : |||:|
 MKVLLIIPAYNEENILRTIASIETFKQEVTHFQHELDYVV
 10 20 30 40

486 516 546 576 606 636 666 696
 INDGSTDGTPELLDRLGLNHI DLVQNLGIGGCVQTYLYANRNHYDVAVQFDGQGHDIRSIEDVVMPIILDEADFVIGS
 ||||| :| :| :| ||||| ||||| ||| :| :| :| :|

-1365-

```

INDGSTDGTKQILEVNQINAIHLVLNLGIGGAVQTGYKYALENEYDVAXQFDGDGXHDIXSLFILLEPLAEGXCXFSXGS
      60      70      80      90     100     110     120

726      756      786      816      846      876      906      936
5 RFVDKKHQNFQSTAMRRIGINLISAAIKLTTGHKVYDTTSGYRAANAALIAVLSCHYPVQYPEPESTARILKKGYRLKEV
  ||:      :|||  |||  || |:|      | :|  ||  ||  :||:::  ||  |||||  ::||  : : |
RFIPGNXASFQSKMRRXGIRLLSFCXXXAXGXTIYKVIXGXRAGNRKVIAFFAKRYPTINYPEPESTIVHLIKKRFVIVER
      140      150      160      170      180      190      200

10 966      996      1026      1056      1086      1116      1146      1176
TANMFEREAGTSSISLKSIFYMTDVLTSIIAGFIKEDDK*V*HCKLKCLF*PLSYFI*L*EWLIKTHFLNLVLYLGY*
  ||  ||  | |||  :|  |:  ||  :|  :::|  ||  ||  |
PVMMERLGGVSSIRALASVKYMLEVGSAILIAPFMKEGD
      220      230      240

15

```

SEQ ID 8752 (GBS355) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 4; MW 27kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 7; MW 52kDa).

GBS355-GST was purified as shown in Figure 213 (lane 4) and in Figure 216 (lane 6).

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1219

A DNA sequence (GBSx1295) was identified in *S.agalactiae* <SEQ ID 3781> which encodes the amino acid sequence <SEQ ID 3782>. Analysis of this protein sequence reveals the following:

```

25 Possible site: 19
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.91    Transmembrane  185 - 201 ( 185 - 201)

30 ----- Final Results -----
        bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

35 >GP:BAA32090 GB:AB010970 rhamnosyltransferase [Streptococcus mutans]
    Identities = 181/315 (57%), Positives = 244/315 (77%), Gaps = 7/315 (2%)

Query: 1 MKVNILMATYNGEKFLAQQIESIQKQTFKEWNLLIRDDGSSDKTCDIIRNFTAKDSRIRF 60
      MKVNILM+TYNG++F+AQQI+SIQKQTF+ WNLLIRDDGSSD T II +F D+RIRF
40 Sbjct: 1 MKVNILMSTYNGQEFIAQQIQSTIQKQTFENWNLLIRDDGSSDGTKIADFAKSDARIRF 60

Query: 61 INENEHHNLGVIKSFFTLVNVEVADFYFFSDQDDVWLPEKLSVSLEAAKHKASDVPLLVY 120
      IN ++ N GVIK+F+TL+ YE AD+YFFSDQDDVWLPEKLSVSLEAAKHKASDVPLLVY
45 Sbjct: 61 INADKRENFVGIKNFYTLTKYEADYFFSDQDDVWLPEKLSVSLEAAKHKASDVPLLVY 120

Query: 121 TDLKVVNQELNQLQDSMIRAQSHHANTTLPELTENTVTGGTMMINHALAEKW-FTPNDI 179
      TDL VV+++L +L DSMI+ QSHHANT+LL ELTENTVTGGTMM+NH LA++W +D+
50 Sbjct: 121 TDLTVVDRDLQVLHDSMIKTQSHHANTSLEELTENTVTGGTMMVNHCLAKQWKQCYDDL 180

Query: 180 LMHDWFLALLAASLGEEIYLDLPTQLYRQHDNNVLGARTMDKRKF-ILREGPKSIFTRYW 238
      +MHDW+LALLAASLG++IYLD T+LYRQH++NVLGART KR K LR P + +YW
55 Sbjct: 181 IMHDWYLALLAASLGKLIYLDETTELYRQHESNVLGARTWSKRLKNWLR--PHRLVKKYW 238

Query: 239 KLIHDSQKQASLIVDKYCDIMTANDELKICFIKIDKQPFMTRLRLWLKYGYSKNQFKHQ 298
      L+ SQ+QAS +++ D+ AN +I+ ++ + Q F+ R++WL +YG++KN+ H
Sbjct: 239 WLVTSSQQQASHLLEL--DLPAANK-AIIRAYVTLDDQSFLNRIKWLKQYGFKNRAFHT 295

Query: 299 VVFKWLIATNYINKR 313

```

-1366-

VFK LI T + +R
 Sbjct: 296 FVFKTLIITKFGYRR 310

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 817> which encodes the amino acid
 5 sequence <SEQ ID 818>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.1980(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 178/314 (56%), Positives = 232/314 (73%), Gaps = 6/314 (1%)
 Query: 1 MKVNIIMATYNGEKFLAQQIESIQKQTFKEWNLIRDDGSSDKTCDIIRNFTAKDSRIRF 60
 M +NIL++TYNGE+FLA+QI+SIQ+QT +W LLIRDDGS+D T DIIR F +D RI++
 Sbjct: 1 MNINILLSTYNGERFLAEQIQSIQRQTVNDWTLLIRDDGSTDGTQDIIRTFVKEDKRIQW 60
 20 Query: 61 INENEHNLGVIKSFFTLVNYEVADFYFFSDQDDVWLPEKLSVS-LEAAKHKASDVPLLV 119
 INE + NLGVIK+F+TL+ ++ AD YFFSDQDD+WL KL V+ LEA KH+ + PLLV
 Sbjct: 61 INEGQTENLGVIKNFYTLKHKQKADVFFSDQDDIWLNDNKLEVTLLLEAQKHEMT-APLLV 119
 25 Query: 120 YTDLKVVNQELNILQDSMIRAQSHHANTTLLPELTENTVTGGTMMINHALAEKWFTPNDI 179
 YTDLKVV Q L + DSMI+ QS HANT+LL ELTENTVTGGTMMI HALAE+W T + +
 Sbjct: 120 YTDLKVVTOHLAVCHDSMIKTQSGHANTSLLELTENTVTGGTMMITHALAEWTTCDGL 179
 30 Query: 180 LMHDWFLALLAASLGEIYLLDPTQLYRQHDNNVLGARTMDKRFKILREGPKSIFTRYWK 239
 LMHDW+LALLA+++G+++YLD+PT+LYRQHD NVLGART KR K P + +YW
 Sbjct: 180 LMHDWYLLALLASAIGKLVYLDIPTELYRQHDANVLGARTWSKRMKNWLT-PHHLVNKYWW 238
 Query: 240 LIHDSQKQASLIVDKYGDIMTANDELELIKCFIKIDKQPFMTRLRLWLKYGYSKNQFKHQV 299
 LI SQKQA L++D + ND EL+ ++ + PF RL L +YG+ KN+ H
 35 Sbjct: 239 LITSSQKQAQLLLDL---PLKPNDELVTAYVSLDMPFTKRLATLKRYGFRKNRIFHTF 295
 Query: 300 VFKNLIATNYYNKR 313
 +F+ L+ T + +R
 40 Sbjct: 296 IFRSLVVTLFGYRR 309

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1220

A DNA sequence (GBSx1296) was identified in *S.agalactiae* <SEQ ID 3783> which encodes the amino
 45 acid sequence <SEQ ID 3784>. This protein is predicted to be rgpAc. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1881(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 9993> which encodes amino acid sequence <SEQ ID 9994> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1367-

>GP:EAA32089 GB:AB010970 rgpAc [Streptococcus mutans]
Identities = 234/362 (64%), Positives = 284/362 (77%)

5 Query: 33 VSELINHQKSFEDIKYHVACLSDEKHHHTFNFADADCFTINPPQLGPARVIAYDIMAINYA 92
+ EL+ +++S + YHVACLS+ + H HF + DCFTI P+LGPARIAYD+MAI YA
Sbjct: 1 MEELVKYKQSQQLTYHVACLSSETDQKHFTYLVGDCFTIKAPKLGPARVIAYDMMAIRYA 60

10 Query: 93 LDLVKTHDLKEPIFYILGNTIGAFIWHFANKIHKVGGLLVYVNPDLGLEWKRKWSRPTQRY 152
L L+K +K PIFYILGNTIGAF+ FA KI ++GG Y+NPDGLEW+RSKWSRP Q Y
Sbjct: 61 LKLIKQKIKHPIFYILGNTIGAFMGPFARKIKRIGGRFYINPDGLEWRRSKWSRPVQAY 120

15 Query: 153 LKYAEKCMTKNADLIISDNIGIENYIQSTYSNVKTRFIAYGTEINSRKLSSDDPRVKQLF 212
LKYAEKCMTK ADL+ISDN GIE YI+ Y KT FIAYGT+++ L +D +VK +
Sbjct: 121 LKYAEKCMTKKADLVISDNTGIEGYIKQMPYAKTTFIAYGTDLSPSGLLKNDKVKDFY 180

20 Query: 213 KKWNISKGYYLIVGRFVPENNYETAIREFMASDTKRDVLIICNHQNNPYFEKLSLKTNL 272
KKW IK KGYYLIVGRFVPENNYETAIREFM S ++RDLVIICN++ N YFE L KT
Sbjct: 181 KKAIDKGYYLIVGRFVPENNYETAIREFMTSSSERDLVIICNYEGNAYFEDLRQKTEF 240

25 Query: 273 QQDKRVKVFVGTLYEKDLLDYVRQQAFAYIHGHEVGGTNPGLEALANTDLNLVLDVDFNK 332
+DKR+KFVGT+Y++ LL Y+R+QAFAYIHGHEVGGTNPGLEALALA+TDLNLVL +FN
Sbjct: 241 DKDKRIKFVGTVDYDRPLTYIREQAFAYIHGHEVGGTNPGLEALAHNTDLNLVLITEFNY 300

30 Query: 333 SVAGLSSFFWAKKEGDLAKLINDSDQQDLSTYGDRAKAIQENYTWKKIVEEYEDLFLN 392
+VA ++ YW + G LA+LIN D+Q++ + YG RAK II YTW+KIVEEYEDLFL+
Sbjct: 301 TVALDAARYWTQDNGSLAQLNQFDKQENFAEYGGRAKETIIVNYTWEKIVEEYEDLFLH 360

Query: 393 ES 394
ES
Sbjct: 361 ES 362

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3785> which encodes the amino acid sequence <SEQ ID 3786>. Analysis of this protein sequence reveals the following:

35 Possible site: 23
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.38 Transmembrane 95 - 111 (95 - 111)

40 ----- Final Results -----
bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 250/383 (65%), Positives = 307/383 (79%)

Query: 11 MQDVFIIGSRGLPARYGGFETVSELIHQKSFEDIKYHVACLSDEKHHHTFNFADADCFT 70
MQDVFIIGSRGLPA+YGGFETV ELI+HQ S +I+YHVACLSD +H HF++ ADCF
Sbjct: 1 MQDVFIIGSRGLPAKYGGFETVVEELISHQSSKNIRYHVACLSDTKHKVHFVYKGDACFY 60

50 Query: 71 INPPQLGPARVIAYDIMAINYALDLVKTHDLKEPIFYILGNTIGAFIWHFANKIHKVGG 130
+NPP+LGPARIAYD+MAI YAL H ++ PIFY+LGNT+GAFI F +IH GG
Sbjct: 61 LNPPKLGPARVIAYDMMAITYALSYSQHQIQNPIFYVLGNTVGAFIAPFVKQIHNRRGR 120

55 Query: 131 LYVNPDLGLEWKRKWSRPTQRYLKYAEKCMTKNADLIISDNIGIENYIQSTYSNVKTRFI 190
++NPDGLEWKRKWSRP Q YLK++EK MT+ ADL+ISDNIGI+ Y++ Y KT FI
Sbjct: 121 FFINPDGLEWKRKWSRPVQAYLKFSKQMTROADLVISDNIGIDRYLKQVYPWSKTCFI 180

60 Query: 191 AYGTEINSRKLSSDDPRVKQLFKKWNISKGYYLIVGRFVPENNYETAIREFMASDTKRD 250
AYGT+ +L++ D +V+ F+ ++I+ K YYLI+GRFVPENNYETAI+EFMAS TKRD
Sbjct: 181 AYGTQTQPSRLATADSKVRAYFTFEDIREKDYLLILGRFVPENNYETAIKEFMASSTKRD 240

65 Query: 251 LVIIICNHQNNPYFEKLSLKTNLQQDKRVKVFVGTLYEKDLLDYVRQQAFAYIHGHEVGGTN 310
LVIIICNH+ N YF++L +T +D R+KFVGTLY+K+LL Y+R+QA+AYIHGHEVGGTN
Sbjct: 241 LVIIICNHEGNAYFKQLLAETECDDPRIKFVGTLYDKELLAYIREQAYAYIHGHEVGGTN 300

-1368-

Query: 311 PGLLEALANTDNLNVLVDVDFNKSVAGLSSFYWAKKEGDLAKLINDSDQQQDLSTYGDRAK 370
 PGLLEALA+T+LNLVL VDFN+SVA ++ YW K++G LA+LIN D D G AK
 Sbjct: 301 PGLLEALAHTNLTNLVLGVDFNQSVAKSAALYWTQKQGLAELINQVDAGFDSHLGKEAK 360

5 Query: 371 AIIQENYTWKKIVEEYEDLFLNE 393
 AIIQE+YTW+KIV EYE LFLNE
 Sbjct: 361 AIIQEHYTWKIVGEYEALFLNE 383

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 1221

A DNA sequence (GBSx1297) was identified in *S.agalactiae* <SEQ ID 3787> which encodes the amino acid sequence <SEQ ID 3788>. This protein is predicted to be dTDP-L-rhamnose synthase. Analysis of this protein sequence reveals the following:

15 Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.1059(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAD10184 GB:AF026471 Cps20 [Streptococcus pneumoniae]
 Identities = 258/283 (91%), Positives = 274/283 (96%)

Query: 1 MILITGANGQLGSELRLDERTQEYVAVDVAEMDITNAEMVDKVFEEVKPSLVYHCAAY 60
 MILITGANGQLG+ELR+LLDER +EYVAVDVAEMDIT+AEMV+KVFEEVKP+LVYHCAAY
 Sbjct: 1 MILITGANGQLGTELRYLLDERNEEYVAVDVAEMDITDAEMVEKVFEEVKPTLVYHCAAY 60

30 Query: 61 TAVDAAEDEGKELDFAINVTGTENVAKAAKHDAITLVYISTDYVFDGEKFPVGQEWVDDL 120
 TAVDAAEDEGKELDFAINVTGT+NVAKA+ KH ATLVYISTDYVFDG+KPVGQEWVDD
 Sbjct: 61 TAVDAAEDEGKELDFAINVTGTKNVAKASEKHGATLVYISTDYVFDGKFPVGQEWVDDR 120

35 Query: 121 PDPKTEYGRTRKMGEEELVEKYTSKYFTIRTAWVFGNYGKNFVFTMQNLAKTHKTLTVVND 180
 PDP+TEYGRTRKMGEEELVEK+ S FY IRTAWVFGNYGKNFVFTMQNLAKTHKTLTVVND
 Sbjct: 121 PDPQTEYGRTRKMGEEELVEKHVSNFYIIRTAWVFGNYGKNFVFTMQNLAKTHKTLTVVND 180

40 Query: 181 QHGRPTWTRTLAEFMTYLAENQKDFGYYHLSNDAKEDTTWYDFAVEILKDTDVEVKPVDS 240
 Q+GRPTWTRTLAEFMTYLAEN+K+FGYYHLSNDA EDTTWYDFAVEILKDTDVEVKPVDS
 Sbjct: 181 QYGRPTWTRTLAEFMTYLAENRKRFGYYHLSNDAEDTTWYDFAVEILKDTDVEVKPVDS 240

Query: 241 SQFPAKAKRPLNSTMSLEKAKATGFVIPTWQDALKEFYKQEVK 283
 SQFPAKAKRPLNSTMSL KAKATGFVIPTWQDAL+EFYKQEV+
 45 Sbjct: 241 SQFPAKAKRPLNSTMSLAKAKATGFVIPTWQDALQEFYKQEV 283

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3789> which encodes the amino acid sequence <SEQ ID 3790>. Analysis of this protein sequence reveals the following:

50 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 55 bacterial cytoplasm --- Certainty=0.0618(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 227/284 (79%), Positives = 248/284 (86%)

-1369-

5 Query: 1 MILITGANGQLGSELRLDERTQEYVAVDVAEMDITNAEMVDKVFEEVKPSLVYHCAAY 60
 MILITG+NGQLG+ELR+LLDER +YVAVDVAEMDITN + V+ VF +VKP+LVYHCAAY
 Sbjct: 21 MILITG+NGQLGTELRLLDERGVDYVAVDVAEMDITNEDKVEAVFAQVKPTLVYHCAAY 80

10 Query: 61 TAVDAAEDEGKELDFAINVTGTENVAKAAKHDTLVYISTDYVFDGEKPVGQEWVDDL 120
 TAVDAAEDEGK L+ AINVTG+EN+AKA K+ ATLVIYSTDYVFDG KPVGQEW D
 Sbjct: 81 TAVDAAEDEGKALNEAINVTGSENIKACGKYGATLVYISTDYVFDGNKFPVGQEWVETDH 140

15 Query: 121 PDPKTEYGRTKRMGEELVEKYTSKFYTIRTAWVFGNYGKNFVFTMQNLAKTHKTLTVVND 180
 PDPKTEYGRTKR+GE VE+Y FY IRTAWVFGNYGKNFVFTM+ LA+ H LTVVND
 Sbjct: 141 PDPKTEYGRTKRLGELAVERYAEHFYIIRTAWVFGNYGKNFVFTMEQLAENHSRLTVVND 200

20 Query: 181 QHGRPTWTRTLAEFMTYLAENQKDFGYHLSNDAKEDTTWYDFAVEILKDTDVEVKPVDS 240
 QHGRPTWTRTLAEFM YL ENQK FGYHLSNDAKEDTTWYDFA EILKD VEV PVDS
 Sbjct: 201 QHGRPTWTRTLAEFM CYLTENQKAFGYHLSNDAKEDTTWYDFAEILKDKAVEVVPVDS 260

Query: 241 SQFPAKAKRPLNSTMSLEKAKATGFVIPTWQDALKEFYKQEVKK 284
 S FPAKAKRPLNSTM+L+KAKATGFVIPTWQ+ALK FY+Q +KK
 Sbjct: 261 SAFPAKAKRPLNSTMNLDKAKATGFVIPTWQEALKAIFYQQLKK 304

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1222

25 A DNA sequence (GBSx1298) was identified in *S. agalactiae* <SEQ ID 3791> which encodes the amino acid sequence <SEQ ID 3792>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2554(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA21508 GB:AB000631 unnamed protein product [Streptococcus mutans]
 Identities = 92/108 (85%), Positives = 100/108 (92%)

40 Query: 5 KOYSEEEVGKIKDRILEALEMVIDPELGIDIVNLGLIYEIRFEDNGRTEIDMTLTMTMGCP 64
 K Y+ EE+ KIKDRILEALEMVIDPELGIDIVNLGLIY+IRFED+GRTEIDMTLTMTMGCP
 Sbjct: 4 KNYTPEEIAKIKDRILEALEMVIDPELGIDIVNLGLIYDIRFEDSGRTEIDMTLTMTMGCP 63

45 Query: 65 LADLLTDQIHDVMKTVPEVTETEVKLVWYPAWSDKMSRYARIALGIR 112
 LADLLTDQIHD +K VPEV + +VKLVW PAW+VDKMSRYARIALGIR
 Sbjct: 64 LADLLTDQIHDALKDVPEVLDIDVKLVWSPAWTVDKMSRYARIALGIR 111

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3793> which encodes the amino acid sequence <SEQ ID 3794>. Analysis of this protein sequence reveals the following:

50 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2818(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 90/112 (80%), Positives = 102/112 (90%)

-1370-

Query: 1 MSEVKQYSEEEVGKIKDRILEALEMVIDPELGIDIVNLGLIYEIRFEDNGRTEIDMTLTT 60
 MS+ +Y+++V IK+RILEALE VIDPELGID+VNLGLIYEIRF DNG TEIDMTLTT
 Sbjct: 1 MSDTPKYTQDQVIAIKNRILEALETVIDPELGIDVVNLGLIYEIRFNDNGYTEIDMTLTT 60

5 Query: 61 MGCPLADLLTDQIHDVMKTVPEVTETEVKLVWYPAWSVDKMSRYARIALGIR 112
 MGCPLADLLTD IHD ++ VPEVT+TEVKLVWYPAW+VDKMSRYARIALGIR
 Sbjct: 61 MGCPLADLLTDYIHDALQDVPEVTETEVKLVWYPAWTVDKMSRYARIALGIR 112

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1223

A DNA sequence (GBSx1299) was identified in *S.galactiae* <SEQ ID 3795> which encodes the amino acid sequence <SEQ ID 3796>. This protein is predicted to be RNA polymerase sigma factor, sigma-70 family (rpoD). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3157(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to the sigma-42 protein from *S.mutans*:

>GP:BAA21507 GB:AB000631 sigma 42 protein [Streptococcus mutans]
 Identities = 345/367 (94%), Positives = 358/367 (97%)

Query: 14 EKKGNTTFNVQVADFIRNHKKQGTADDEVTEKLVIPFVLADQIDDLRLTDGGISIT 73
 +KK ++TFNVQVADFIRNHKK+G A+DDEVTEKLVIPF L+A+QIDDLRLTDGGISIT
 Sbjct: 5 KKKTSTTFNVQVADFIRNHKKEGVAVDDEVTEKLVIPFELEAEQIDDLRLTDGGISIT 64

Query: 74 DKEGNPSTKYVVEGPKPEELTDEELIGSNSAKVNDPVRMYLKEIGVPLLTNEEEKELAV 133
 D+EGNPSTKY VE KPEELTDEEL+GSNSAKVNDPVRMYLKEIGVPLLTNEEEKELA+
 Sbjct: 65 DREGNPSTKYAVEEIKPEELTDEELIGSNSAKVNDPVRMYLKEIGVPLLTNEEEKELAI 124

Query: 134 AVAEGDLMAKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKGFK 193
 AV GDL AKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKGFK
 Sbjct: 125 AVENGDLAKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKGFK 184

Query: 194 FSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIAER 253
 FSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIAER
 Sbjct: 185 FSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIAER 244

Query: 254 MDMPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIEPNVDYTTTRVVLREQLDE 313
 MDMPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIEPNVDYTTTRVVLREQLDE
 Sbjct: 245 MDMPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIEPNVDYTTTRVVLREQLDE 304

Query: 314 VLDTLTDREENVLRRLRFLDDGKMRITLEDVGKVFVTRERIRQIEAKALRKL RHPSRSKQ 373
 VLDTLTDREENVLRRLRFLDDGKMRITLEDVGKVF+VTRERIRQIEAKALRKL RHPSRSKQ
 Sbjct: 305 VLDTLTDREENVLRRLRFLDDGKMRITLEDVGKVFVTRERIRQIEAKALRKL RHPSRSKQ 364

Query: 374 LKDFMED 380
 L+DF+ED
 Sbjct: 365 LRDFVED 371

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3797> which encodes the amino acid sequence <SEQ ID 3798>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

-1371-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1788(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 351/369 (95%), Positives = 364/369 (98%)

10 Query: 12 MAEKKGNTTFNVQVADFIRNHKKQGTAIIDDEVTEKLVIPFVLADQIDDLERLTDGGIS 71
 M ++K TTFNVQVA+FIR+HKK+GTAIDD+VTEKLVIPF LDADQIDDLERLTDGGIS
 Sbjct: 1 MTKQKEITTFNVQVAEFIRHHKKEGTAIDDDVTEKLVIPFALDADQIDDLERLTDGGIS 60

15 Query: 72 ITDKEGNPSTKYVVEGPKPEELTDEELIGSNSAKVNDPVRMYLKEIGVVPLLTNEEEKEL 131
 ITDKEGNPS+KY+VE PKPEELTDEELIGSNSAKVNDPVRMYLKEIGVVPLLT+EEEEKEL
 Sbjct: 61 ITDKEGNPSSKYIVEEPKPEELTDEELIGSNSAKVNDPVRMYLKEIGVVPLLTSEEEKEL 120

20 Query: 132 AVAVAEGDLMAKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKG 191
 AVAVA+GDLMAKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKG
 Sbjct: 121 AVAVAKGDLMAKQRLAEANLRLVVSIAKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKG 180

25 Query: 192 FKFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIA 251
 FKFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIA
 Sbjct: 181 FKFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQIA 240

30 Query: 252 ERMDMTPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVENPVDYTTRVVLREQ 311
 ERM+MTPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVENPVDYTTRVVLREQ
 Sbjct: 241 ERMEMTPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVENPVDYTTRVVLREQ 300

35 Query: 312 DEVLDTLTDREENVLRRLRFGLDDGKMRTLEDVGKVFNVTRERIRQIEAKALRKL RHPSRS 371
 DEVLDTLTDREENVLRRLRFGLDDGKMRTLEDVGKVFNVTRERIRQIEAKALRKL RHPSRS
 Sbjct: 301 DEVLDTLTDREENVLRRLRFGLDDGKMRTLEDVGKVFNVTRERIRQIEAKALRKL RHPSRS 360

Query: 372 KQLKDFMED 380
 KQL+DF+ED
 Sbjct: 361 KQLRDFIED 369

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1224

40 A DNA sequence (GBSx1300) was identified in *S.agalactiae* <SEQ ID 3799> which encodes the amino acid sequence <SEQ ID 3800>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

45

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2853(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1372-

Example 1225

A DNA sequence (GBSx1301) was identified in *S.agalactiae* <SEQ ID 3801> which encodes the amino acid sequence <SEQ ID 3802>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2198(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA03516 GB:D14690 DNA primase [Lactococcus lactis]
Identities = 206/398 (51%), Positives = 294/398 (73%), Gaps = 6/398 (1%)
15
Query: 37 LAIDKEKISEIKNSVNIVDVIGEVVGLTKTGRNHLGLCPFHKEKTPSFNVIEDRQFFHCF 96
      +++D E ++++K+ VNI D+I + V L++TG+N++GLCPFH EKTSPSFNV ++ F+HCF
Sbjct: 2 VSLDTEVVNDLKSQVNIADLISQYVALSRTGKNYIGLCPFHGEKTPSFNVNAEKGFFHCF 61

Query: 97 GCGRSGDVFKFVEDYQHISFLDSVQVLAERSGIPLDTNFKGQVPPKPKANQSILDIHRVA 156
      GCGRSGD +F+++Y + F+D+V+ LA+ +G+ L N +K N L +I+ A
Sbjct: 62 GCGRSGDAIEFLKEYNQVGFVDAVKELADFAGVTL--NISDDREKNNPNAPLFEINNQA 119

Query: 157 SGFYHAYLMTINDGERARQVLAERGVTEDLIKHFQIGLSPGGQDFLYRRLAKEFDEKTL 216
      + Y+ LM+T GERAR+YL ERG+T+D+IK F IGL+P DF+++ L+ +FDE+ +
25 Sbjct: 120 ARLYNILLMSTELGERARKYLEERGITDDVIKRFNIGLAPEENDFIFKNLSNKFDEEIMA 179

Query: 217 SSGLFNYSSENSNQFYDSFNNRIMFPLTNDIGEVIAPSGRVWTQEDIDRKQAKYKNSRATP 276
      SGLF++S +N+ +D+F NRIMFP+TN+ G+ I FSGR W QE+ D K AKY N+ AT
30 Sbjct: 180 KSGLFPHFS--NNKVFDATFNRMFPITNEYGQTIGFSGRW--QENDDSK--AKYINTSATT 235

Query: 277 IFNKSYLEYHLDKARAVINKAHEVYLMEGFMDVIAAYRAGIENVVASMGTALTNEHVRHL 336
      IF+KSYEL++LDKA+ I+K HEVYLMEGFMDVIA+Y+AGI NVVASMGTALT +HVR L
Sbjct: 236 IFDKSYELWNLDKAKPTISKQHEVYLMEGFMDVIAASYKAGINNVVASMGTALTEKHVRRL 295

Query: 337 KRFTKKVVLTYDGDRAQNAIDKSLELLSDMTVDIVRIPNKMDPDEFLQANSAEDFKQLL 396
      K+ KK VL YGDG AQONAI K+++L+ + V IV++P +DPDE+ + + L+
35 Sbjct: 296 QMAKKFVLVYDGD SAGONAIYKAIDLIGESAVQIVKVPGLDPDEYSKNYGLKGLSALM 355

Query: 397 ENGRISNTEFYIHYLKPENTDNLQSEIAYVEKIAKLIA 434
      E GRI EF I YL+PEN NLQ+++ ++E+I+ +IA
40 Sbjct: 356 ETGRIQPIEFLIDYLRPENLANLQTLQDFIEQISPMIA 393

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3803> which encodes the amino acid sequence <SEQ ID 3804>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3532(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 378/604 (62%), Positives = 477/604 (78%), Gaps = 2/604 (0%)
55
Query: 28 MGYFCGGHDLAIDKEKISEIKNSVNIVDVIGEVVGLTKTGRNHLGLCPFHKEKTPSFNVI 87
      MG+ GG DLAIDKE IS++KNSVNIVDVIGEVV L+++GR++LGLCPFHKEKTPSFNV+
Sbjct: 1 MGFLWGGDDLAIKEMISQVKNVNIVDVIGEVVKLRSRGRHYLGLCPFHKEKTPSFNVV 60

Query: 88 EDRQFFHCFGCGRSGDVFKFVEDYQHISFLDSVQVLAERSGIPLDTNFKGQV--PKKPKA 145
60

```

-1373-

EDRQFFHCFGCG+SGDVFKF+E+Y+ + FL+SVQ++A+++G+ L+ V +
 5 Sb|ct: 61 EDRQFFHCFGCGKSGDVFKFIEEYRQVPFLESVQIIADKTGMSLNIPPSQAVLASQHKHP 120

Query: 146 NQSLLDIHRVASGFYHAYLMTTNDGERARQYLAERGVTEDLIKHFQIGLSPGGQDFLYRR 205
 N +L+ +H A+ FYHA LMTT G+ AR+YL +RG+ + LI+HF IGL+P D+LY+
 10 Sb|ct: 121 NHALMTLHEDAAKFYHAYLMTTTTIGQEARKYLYQRLDDQLIEHFNIGLAPDES DYLYQA 180

Query: 206 LAKEFDEKTLMSGLFNYSSENSNQFYDSFNRMIFPLINDIGEVIASFGRVWTQEDIDRK 265
 L+K+++E L++SGLF+ S+ SN YD+F NRIMFPL++D G +IAFSGR+WT D++++
 10 Sb|ct: 181 LSKKYEEGQLVASGLFHLSDQSNITTYDAFRNRIMFPLSDDRGHIIAFSGRIWTAADMEKR 240

Query: 266 QAKYKNSRATPIFNKSYELYHLDKARAVINKAHEVYLMEGFMDVIAAYRAGIENVVASM 325
 QAKYKNSR T +FNKSYELYHLDKAR VI K HEV+LMEGFMDVIAAYR+G EN VASM
 15 Sb|ct: 241 QAKYKNSRGTVLEFNKSYELYHLDKARPVIKTHEVFLMEGFMDVIAAYRSGYENAVASM 300

Query: 326 TALTNEHVRHLKRFTKKVVLTYDGDRAQNAIDKSLELLSDMTVDIVRIPNKMDPDEF 385
 TALT EHV HLK+ TKKVVL YDGD AGQ+AI KSELL D V+IVRIPNKMDPDEF+Q
 20 Sb|ct: 301 TALTQEHVNHKQVTKKVVLTYDGDAGQHAIAKSELLKDFVVEIVRIPNKMDPDEFVQ 360

Query: 386 ANSAEDFKQLLENGRISNTEFYIHYLKPENTDNLQSEIAYVEKIAKLIKSPSITAQNSY 445
 +S E F LL+ RIS+ EF+I YLKP N DNLSQ+I YVEK+A LIA+SPSITAQ+SY
 25 Sb|ct: 361 RHPSEAFADLLKQSRISSEFFIDYLPKPTNVDNLQSQIVYVEKMAPLIAQSPSITAQHSY 420

Query: 446 ITKVAELLPDFDYFQVEQSVNNERLHRSQQQASSSVQTSATVQLPQTGKLSAITKTEM 505
 I K+A+LLP+FDYFQVEQSVN R+ R + Q + S V LP L+AI KTE
 25 Sb|ct: 421 INKIADLLPNFDYFQVEQSVNALRIQDRQKHQGIQAQAVSNLVTLPMPKSLTAIAKTESH 480

Query: 506 LFHRLLNHPYLLNEFRNRDNFYFDTTEIQVLYELLKESGEITSYDLSQESDKVNRYYII 565
 L HRL+H YLLNEFR+RD+FYFDT+ +++LY+ LK+ G ITSYDLS+ S++VNR YY +
 30 Sb|ct: 481 LMHRLHHDYLLNEFRHRDDFYFDTSTLELLYQRLKQGHITSYDLSSEMSEEVNRAYNV 540

Query: 566 LEEQLPVEVSIGEIEAVEKARDRLKRDRLKQSQLIRQSSNQDEEGALAALENLIAQK 625
 LEE LP EV++GEI+ + R +LL ERDL KQ + +R+SSN+GD + AL LE+ IAQK
 35 Sb|ct: 541 LEENLPKEVALGEIDDILSKRAKLLAERDLHKQGGKVVRESSNKGHDQAALVLEHFTAQK 600

Query: 626 RNME 629
 R ME
 Sb|ct: 601 RKME 604

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1226

A DNA sequence (GBSx1302) was identified in *S. agalactiae* <SEQ ID 3805> which encodes the amino acid sequence <SEQ ID 3806>. Analysis of this protein sequence reveals the following:

45 Possible site: 47
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.05	Transmembrane	41 - 57 (34 - 58)
INTEGRAL	Likelihood = -5.79	Transmembrane	93 - 109 (90 - 112)

50 ----- Final Results -----

bacterial membrane	---	Certainty=0.3421(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

55 A related GBS nucleic acid sequence <SEQ ID 9995> which encodes amino acid sequence <SEQ ID 9996> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC38560 GB:AF029731 large conductance mechanosensitive channel
 [Staphylococcus aureus]
 60 Identities = 64/126 (50%), Positives = 83/126 (65%), Gaps = 8/126 (6%)

-1374-

Query: 23 MIKELKEFLFKGNVLDLAVAVILGAAFNAITTSLVKDVITPLIILNPVLKAAGVSNIA-QL 81
 M+KE KEF KGNVLDLA+AV++GAAFN II+SLV++I PLI K G + A +
 5 Sbjct: 1 MLKEFKFALKGNVLDLATAVVMGAFAFNKIISLVENIIMPLI----GKIFGSVDFAKEW 56

Query: 82 SWNGVAYGNFLSAVINFLIVGTTLFFIVKAANKVMAKKPAEEIIIEVVEPTQEQLLAEIR 141
 S+ G+ YG F+ +VI+F+I+ LF VK AN +M K+ AEE E V LL EIR
 10 Sbjct: 57 SFWGIKYGLFTQSVDFIIIAFALFIFVKIANTLMKKEAE---EAVVEENVVLLTEIR 113

Query: 142 DLLANK 147
 DLL K
 10 Sbjct: 114 DLLREK 119

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3807> which encodes the amino acid sequence <SEQ ID 3808>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -5.95 Transmembrane 71 - 87 (67 - 90)
 20 ----- Final Results -----
 bacterial membrane --- Certainty=0.3378(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GF:CAB15653 GB:Z99122 similar to large conductance mechanosensitive
 channel protein [Bacillus subtilis]
 Identities = 61/126 (48%), Positives = 77/126 (60%), Gaps = 7/126 (5%)
 30 Query: 1 MVKELKAFLEFRGNIIELAVAVIIGGAFGAIVTSFVNDIITPLIILNPALKAAANVENITQLS 60
 M E KAF RGNI++LA+ V+IGGAFG IVTS VNDII PL+ L + ++
 Sbjct: 1 MWNEFKAFAMRGNIIVDLAIGVVIGGAFGKIVTSFVNDIIMPLV-GLLLGGLDFSGLSFTF 59

Query: 61 WNG-VKYGSFLGAVINFLIIGTSLFFVVKAAEKAMPKKE-----KEAAAPTQEELLTEIR 114
 + VKYGSF+ ++NFLII S+F V++ KKE E A QEELL EIR
 35 Sbjct: 60 GDAVVKYGSFIQTIVNFLIISFSIFIVIRTLNGLRRKKEAEAEAEAEVDAQEELLKEIR 119

Query: 115 DLLAQK 120
 DLL Q+
 40 Sbjct: 120 DLLKQQ 125

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/125 (68%), Positives = 99/125 (78%), Gaps = 5/125 (4%)
 45 Query: 23 MIKELKEFLFKGNVLDLAVAVILGAAFNAITTSLVKDVITPLIILNPVLKAAGVSNIAQLS 82
 M+KELK FLF+GN+++LAVAVI+G AF AI+TS V D+ITPLIILNP LKAA V NI QLS
 Sbjct: 1 MVKELKAFLEFRGNIIELAVAVIIGGAFGAIVTSFVNDIITPLIILNPALKAAANVENITQLS 60

Query: 83 WNGVAYGNFLSAVINFLIVGTTLFFIVKAANKVMAKKPAEEIIIEVVEPTQEQLLAEIRD 142
 WNGV YG+FL AVINFLI+GT+LFF+VKAA K M KK E PTQE+LL EIRD
 50 Sbjct: 61 WNGVKYGSFLGAVINFLIIGTSLFFVVKAAEKAMPKKEK-----EAAAPTQEELLTEIRD 115

Query: 143 LLANK 147
 LLA K
 55 Sbjct: 116 LLAQK 120

A related GBS gene <SEQ ID 8753> and protein <SEQ ID 8754> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 60 SRCFLG: 0
 McG: Length of UR: 4
 Peak Value of UR: 2.96

-1376-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9391> which encodes amino acid sequence <SEQ ID 9392> was also identified. A related GBS nucleic acid sequence <SEQ ID 10799> which encodes amino acid sequence <SEQ ID 10800> was also identified.

The protein is similar to the 30S ribosomal protein S21 from *Listeria monocytogenes*:

>GP:BAA82793 GB:AB023064 30S ribosomal protein S21 [Listeria monocytogenes]
Identities = 30/34 (88%), Positives = 34/34 (99%)

10 Query: 1 MTKAGTLQESRKREFYEKPSVKRKRKSEAARKRK 34
++K+GTLQESRKREFYEKPSVKRKRKSEAARKRK
Sbjct: 23 VSKSGTLQESRKREFYEKPSVKRKRKSEAARKRK 56

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3811> which encodes the amino acid sequence <SEQ ID 3812>. Analysis of this protein sequence reveals the following:

Possible site: 38
>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4815(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 35/36 (97%), Positives = 36/36 (99%)
Query: 1 MTKAGTLQESRKREFYEKPSVKRKRKSEAARKRKKF 36
+TKAGTLQESRKREFYEKPSVKRKRKSEAARKRKKF
30 Sbjct: 35 VTKAGTLQESRKREFYEKPSVKRKRKSEAARKRKKF 70

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1228

A DNA sequence (GBSx1304) was identified in *S.agalactiae* <SEQ ID 3813> which encodes the amino acid sequence <SEQ ID 3814>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -7.06 Transmembrane 5 - 21 (3 - 23)
40 INTEGRAL Likelihood = -2.28 Transmembrane 191 - 207 (189 - 207)
----- Final Results -----
bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8755> and protein <SEQ ID 8756> were also identified. Analysis of this protein sequence reveals the following:

50 Lipop Possible site: -1 Crend: 2
McG: Discrim Score: 8.68
GvH: Signal Score (-7.5): -5.71

-1377-

Possible site: 18
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 2 value: -7.06 threshold: 0.0
 INTEGRAL Likelihood = -7.06 Transmembrane 5 - 21 (3 - 23)
 INTEGRAL Likelihood = -2.28 Transmembrane 191 - 207 (189 - 207)
 PERIPHERAL Likelihood = 4.35 142
 modified ALOM score: 1.91

*** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.3824(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

SEQ ID 8756 (GBS259) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 4; MW 54kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 1229

A DNA sequence (GBSx1305) was identified in *S.agalactiae* <SEQ ID 3815> which encodes the amino acid sequence <SEQ ID 3816>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.38 Transmembrane 136 - 152 (135 - 152)

----- Final Results -----
 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD47593 GB:AF140784 Vexp2 [Streptococcus pneumoniae]
 Identities = 117/212 (55%), Positives = 152/212 (71%)

Query: 1 MLELKNIAYRYKGNNDKNTLENINYSFQSGVFYITILGNSGSGKTTLLSLMAGLDSPTGQV 60
 +L+L+++ YRYK L INY+F+ G FY+I+G SG+GK+TLLSL+AGLDSP EG +
 Sbjct: 3 LLQLQDVTRYKNTAEAVLYQINYNFEPGKIFYSIIGESGAGKSTLLSLLAGLDSPEVGS 62

Query: 61 LFNKKDIKEAGYAQHRKKNIALVFQNYNLLDYLTPLENVQLVKPTADKQLLLDLGLKEDM 120
 LF +DI++ GY+ HR +I+LVFQNYNL+DYL+PLEN++LV A K LL+LGL E
 Sbjct: 63 LFQGEDIRKKGYSYHRMHHSILVFQNYNLIDYLSPLENIRLVNKKASKNTLLELGLDESQ 122

Query: 121 LTRNLRSLSGGQQQRVAIARALVVGTPAILLDEPTGNLDFDISRDITMRLKDFAHKEKRC 180
 + RN+L+LSGGQQQRVAIAR+LV P IL DEPTGNLD + DI LK A K +C
 Sbjct: 123 IKRNVLQLSGGQQQRVAIARSLVSEAPVILADEPTGNLDEKTAGDIVELLKSLAQKTGKC 182

Query: 181 VIMVTHSREIAHMADTALQLIGDNLKELSKES 212
 VI+VTHS+E+A +D L+L L E S
 Sbjct: 183 VIVVTHSKEVAQASDITLELKDKKLTETRNTS 214

SEQ ID 3816 (GBS363) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 5; MW 28kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 10; MW 53kDa).

GBS363-GST was purified as shown in Figure 216, lane 9.

-1378-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1230

A DNA sequence (GBSx1306) was identified in *S.agalactiae* <SEQ ID 3817> which encodes the amino acid sequence <SEQ ID 3818>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have an uncleavable N-term signal seq
10  INTEGRAL    Likelihood =-14.97  Transmembrane    71 - 87 ( 66 - 97)
    INTEGRAL    Likelihood = -3.61  Transmembrane     2 - 18 ( 1 - 18)

----- Final Results -----
        bacterial membrane --- Certainty=0.6986(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1231

A DNA sequence (GBSx1307) was identified in *S.agalactiae* <SEQ ID 3819> which encodes the amino acid sequence <SEQ ID 3820>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

```

25  Possible site: 45
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
        bacterial cytoplasm --- Certainty=0.1986(Affirmative) < succ>
30        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1232

A DNA sequence (GBSx1308) was identified in *S.agalactiae* <SEQ ID 3821> which encodes the amino acid sequence <SEQ ID 3822>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

```

40  Possible site: 34
    >>> Seems to have no N-terminal signal sequence
        INTEGRAL    Likelihood = -6.05  Transmembrane    22 - 38 ( 17 - 39)

45  ----- Final Results -----
        bacterial membrane --- Certainty=0.3421(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```


-1379-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAD47594 GB:AF140784 Vexp3 [Streptococcus pneumoniae]
 Identities = 39/153 (25%), Positives = 67/153 (43%), Gaps = 9/153 (5%)

Query: 3 LFKRSFLYVSRKKRSITLFLVCLWLVASTLISGIAVKNAGLTA-KKTFSRQTGSILHISS 61
 + +F YV+RK KSI +F+ + L+AS + G+++K A A ++TF T S +
 10 Sbjct: 1 MLHNAPAYVTRKFFKSIVIFLIILLMASLSLVGLSIKGATAKASQETFKNITNS-FSMQI 59

Query: 62 DSTDLVGDGYGSGEIPKAIIVNIASNPVNRVNNLMAYAGLTSEKMVTRPNDKEQYKE- 120
 + G G+G I + I I N ++ + A LT ++ P K+
 15 Sbjct: 60 NRRVNQGTTPRGAGNIKGEDIKKITENKAIESYVVKRINAIGDLTGYDLIETPETKKNLTAD 119

Query: 121 -----QVLQVHGNSYSDTDPKYTAGMISLKGK 147
 L + G + S + K+ +G L G
 15 Sbjct: 120 RAKRFGSSSLMITGVNDSSKEDKVFVSGSYKLVEG 152

No corresponding DNA sequence was identified in *S.pyogenes*.

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1233

A DNA sequence (GBSx1309) was identified in *S.agalactiae* <SEQ ID 3823> which encodes the amino acid sequence <SEQ ID 3824>. Analysis of this protein sequence reveals the following:

25 Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -15.76	Transmembrane	295 - 311 (287 - 317)
INTEGRAL	Likelihood = -7.59	Transmembrane	49 - 65 (46 - 69)
INTEGRAL	Likelihood = -6.90	Transmembrane	340 - 356 (339 - 362)
INTEGRAL	Likelihood = -5.57	Transmembrane	411 - 427 (404 - 430)

30 ----- Final Results -----

bacterial membrane	---	Certainty=0.7305 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

35

A related GBS nucleic acid sequence <SEQ ID 9695> which encodes amino acid sequence <SEQ ID 9696> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB12182 GB:Z99106 similar to transporter [Bacillus subtilis]
 Identities = 95/370 (25%), Positives = 167/370 (44%), Gaps = 41/370 (11%)

Query: 109 ESVEASLSIDVGSRLKSVSPYNSS-----KBENQVTLAGYQSTEDLRAFQTKALVLK 160
 +++E+S S D S S + NS + +++ G ST + F +
 45 Sbjct: 115 DAIESSSSSSSSSSSSSNAKNSQGGGQGGPQMVQADLSIEGVISTALVDDFSDGDSKIT 174

Query: 161 KGSHLAADNT--KQVLVPLKLAQKNHLSVGNKLRGK---ENVT----IAGIYDANSA-- 209
 G + + K ++ LA++N LSVG+ + + E+ T I GIY S+
 50 Sbjct: 175 DGRAITKSDVGKKTVINETLAEENDLSVGDSITIESATDEDTTVKLKIVGIYKTTSSGD 234

Query: 210 -KSKNTFNPNIIDNTLIAQATLVKRKISKQKGQTV--AVRLSDKRLVDTVIONIKQWPLD 265
 +++N N N L T + T+ + D + +DT ++ K+ +D
 55 Sbjct: 235 DQAQNFSLNPNKLYTPYTATAALKGDDYKNTIDSAYVYMDAKNMDTFVKAAKKTSID 294

Query: 266 FGKLDVQTAKEFYGDSYRNIEHLRLVGRILLIVSLVAMAILVVMLTFWINNRIKETGIL 325
 F + T + Y IE + ++ +VS+ IL +++ I R E G+L
 60 Sbjct: 295 FDTYTLNTNDQLYQQMVGPENVASFSKNVVYLVSVAGAVILGLIVMMSIRERKYEMGV 354

-1380-

Query: 326 LAIGTKFEIIGHYLIEVLLVAGAAFTLSIIGGVFLGKTFAGLLSQV----- 373
 +AIG+ ++++IG +L E+L+VA A L+ + G + LLSQ
 Sbjct: 355 MAIGEKRWKLIGQFLTEILIVAVIAIGLASVTGNLVANQLGNQLLSQQISSSTDSTQTAS 414

5 Query: 374 -----NGGVSSQIVQNSSLIIDRIDNLAVSVGVMDVFRLYAQGALICLFAVVLSSYSIL 427
 GG+ ++ +SS +D ID+L V+V + D+ L G LI + A +L S S+L
 Sbjct: 415 GQMPGGGGGGMGGKMFHSSSNVDVIDSLNVAVSMNDMLILGGIGILIAIATLLPSISVL 474

10 Query: 428 KLQPKQILSR 437
 +L PK IL++
 Sbjct: 475 RLHPKTIITK 484

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8757> and protein <SEQ ID 8758> were also identified. Analysis of this
 15 protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
 McG: Discrim Score: 1.50
 GvH: Signal Score (-7.5): -8.43
 Possible site: 39

20 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 4 value: -15.76 threshold: 0.0

INTEGRAL	Likelihood = -15.76	Transmembrane	295 - 311 (287 - 317)
INTEGRAL	Likelihood = -7.59	Transmembrane	49 - 65 (46 - 69)
INTEGRAL	Likelihood = -6.90	Transmembrane	340 - 356 (339 - 362)
INTEGRAL	Likelihood = -5.57	Transmembrane	411 - 427 (404 - 430)
PERIPHERAL	Likelihood = 3.45		386

modified ALOM score: 3.65

30 *** Reasoning Step: 3

----- Final Results -----

bacterial membrane	---	Certainty=0.7305(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

ORF00687(421 - 1611 of 1917)
 EGAD|108957|BS0375(11 - 484 of 486) hypothetical protein {Bacillus subtilis}
 OMNI|NT01BS0429 membrane transport protein GP|1805444|dbj|BAA09006.1||D50453 homologue of
 40 hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus
 {Bacillus subtilis} GP|2632675|emb|CAH12182.1||Z99106 similar to transporter {Bacillus
 subtilis} PIR|F69762|F69762 transporter homolog yclI - Bacillus subtilis
 %Match = 8.6
 %Identity = 28.7 %Similarity = 52.2
 45 Matches = 117 Mismatches = 184 Conservative Sub.s = 96

312	342	372	402	432	462	492	522
VL*NH*LIDNVEVDREYLTTSIVILEIIEIKIEGGKIVNLWTLRLAYLKRQKMKTIVTLFLVFLTIGTCLISLMSIQHSLEK							
				:	:	:	:
				:	:	:	:
MNFIKRAFWNMKAKKGTLLQLFVFTVICVFLSGLAIQSAQK							
				10	20	30	40

543	573	603	624	654
N---ILTKQCKSIYLTSSKEKAYWPEQAYEALKK-----AKMVESVEASLSID				
:	:	:	:	:
:	:	:	:	:
SSELARQELGGSVTLQVDRQKQMEKQDQSGEKRTFESTPIKVS DANKLAALDHVKS YNYTTSASANAGNFDAIESSSSSD				
60	70	80	90	100
60	70	80	90	100
110	120			

684	720	750	780	807	834	864
VGSRLKSVSPYNSS-----KEENQVTLAGYQSTEDLRAFQTKALVLKKGSHLA-ADNTKQV-LVPLKLAQKNHLSVG						
:	:	:	:	:	:	:
:	:	:	:	:	:	:
SSSSSSSSNAKNSQGGQGGPQMVQADLSIEGVISTALVDDFSDGDSKITDGRAITKSDVGKKTVINETLAEENDLSVG						
140	150	160	170	180	190	200

-1381-

```

885      903      954      978      1008      1065
NKRLR---GKENVTI---AGIYDANSA---KSKNTFNPNDNTLIA--QATLVRKISKQKGYQTVAVR-LSDKRLVDTV
: : : | : | | | : : : | | | | | | | | : | : : |
DSITIESATDEDTTVKLKIVGIYKTTSSGDDQAQNFSLNPNKLYTPYTATAALKGDDYKNTIDSAVYYMDDAKNMDTF
5      220      230      240      250      260      270      280

1095      1125      1155      1185      1215      1245      1275      1305
IQNIKQWPLDFGKLDVQTAKEFYGDSYRNIETLHRLVGRITLIVSLVAMAILVVMLTFWINNRIKETGILLAIGKTKFEI
: : | : : | | : | : : | | : : : : | | : : : | | : | : | : : :
10 VKAAKTSIDFDTYTLNTNDQLYQQMVGPIENVASFKNVYLVSVAGAVILGLIVMMSIRERKYEMGVLMAIGEKRWKL
300      310      320      330      340      350      360

1335      1365      1395      1431      1461      1491
IGHYLIEVLLVAGAAFTLSIIGGVFLGKTFAAGLLSQV-----NGGVSSQIVQNSSLIIDRIDNLAV
| : : | : : | | | : : | : : : | | | | | : : : | : | : | : |
15 IGQFLTEILIVAVIAIGLASVTGNLVANQLGNQLLSQQISSSTDSTQTASGQMPGGGGMGGMFGHSSSNVDVIDSLNV
380      390      400      410      420      430      440

1521      1551      1581      1611      1641      1671      1701      1731
SVGVMDVFRLYAQGALICLFAVVLSSYSILKLPKQILSRMS*EVNMNLFKRSFLYVSRKKRKSITLFCVLWLVLASTLIS
: | : : : | | | : | : | | : : : | : : :
20 AVSMNDMLILGGIGILIAIATLLPSISVLRHLHPKPTILTKQE
460      470      480

```

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1234

A DNA sequence (GBSx1310) was identified in *S.agalactiae* <SEQ ID 3825> which encodes the amino acid sequence <SEQ ID 3826>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
35 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

40 >GP:CAB11993 GB:Z99105 ybdG [Bacillus subtilis]
Identities = 66/224 (29%), Positives = 102/224 (45%), Gaps = 22/224 (9%)

Query: 84 IKEYGQKVEVKGKKMNVTYVGEQKVPVIFIPQGTVTAKHQYHNLISNLSKTHKVVVVEP 143
+K G V+V GKKMNVTY G GK VF+ G G ++ L S SK +K+ VV+
45 Sbjct: 41 LKGKGTVDVDGKKMNVTYQEGSGKDTFVFMMSGGIAAPAYEMKGLYSKFSKENKIAVVDR 100

Query: 144 FGSLSDVIDQPRNLANITSIDIHEALQKVGITGKYVIASHSIGGVYALKYISTYPKEVLG 203
G G S+V R++ + +AL K G Y++ HSI G+ A+ + YPKE+
50 Sbjct: 101 AGYGYSEVSHDDRDRDITVLEQTRKALMKSGNKPPYILMPHSISGIEAMYWAQKYPKEIKA 160

Query: 204 LIGLDTSTP-----GMEGGKQVDF-----AAPVLKELPKIPKVSDDIN 241
+I +D P G++ K F +A E+ + ++D+
55 Sbjct: 161 IIAMDIGLPQQYVITYKLSGVDRLKVRGFHLLTSIGFHRFIPSAVYNPEVIRQSFLTDEEK 220

Query: 242 AQFFAIGHKILNNSNMKEEAKNSSNMINESANYKIPKGI PAMYL 285
+ AI K N++M+ E S ++S N PK P + L
Sbjct: 221 EIYKAINFKQFFNADMEHELLQSYQNGSKSVNLPAPKETPVLLIL 264

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-1382-

SEQ ID 3826 (GBS121) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 24 (lane 9; MW 40kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 6; MW 65kDa).

GBS121-GST was purified as shown in Figure 198, lane 6.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1235

A DNA sequence (GBSx1311) was identified in *S.agalactiae* <SEQ ID 3827> which encodes the amino acid sequence <SEQ ID 3828>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 33
    >>> Seems to have a cleavable N-term signal seq.

    ----- Final Results -----
    bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
15    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8759> which encodes amino acid sequence <SEQ ID 8760> was also identified. Analysis of this protein sequence reveals the following:

```

20 Lipop: Possible site: -1   Crend: 8
    McG: Discrim Score:      3.70
    GvH: Signal Score (-7.5): -0.0600004
    Possible site: 22
    >>> Seems to have a cleavable N-term signal seq.
25 ALOM program   count: 0 value:  8.01 threshold:  0.0
    PERIPHERAL Likelihood =  8.01      167
    modified ALOM score: -2.10

    *** Reasoning Step: 3
30 ----- Final Results -----
    bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
35    bacterial cytoplasm --- Certainty=0.0000 (Not Clear)

```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8760 (GBS60) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 7; MW 38.6kDa).

GBS60-His was purified as shown in Figure 193, lane 3.

- 40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1236

A DNA sequence (GBSx1312) was identified in *S.agalactiae* <SEQ ID 3829> which encodes the amino acid sequence <SEQ ID 3830>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```

45 Possible site: 21
    >>> May be a lipoprotein

```

5

A related GBS nucleic acid sequence <SEQ ID 9693> which encodes amino acid sequence <SEQ ID 9694> was also identified.

10

15

20

30

35

40

45

50

55

SEQ ID 8762 (GBS21) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 3; MW 31.6kDa).

60 GBS21-His was purified as shown in Figure 192, lane 11.

-1384-

GBS21L was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 124 (lane 8-10; MW 66.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 124 (lane 11; MW 41.5kDa) and in Figure 180 (lane 6; MW 41kDa). GBS21L-His was purified as shown in Figure 232 (lanes 3 & 4)

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1237

- A DNA sequence (GBSx1313) was identified in *S.galactiae* <SEQ ID 3831> which encodes the amino acid sequence <SEQ ID 3832>. This protein is predicted to be endopeptidase O. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3854(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 20 >GP:AAF67832 GB:AF179267 endopeptidase PepO2 [Lactococcus lactis]
Identities = 21/36 (58%), Positives = 26/36 (71%)

Query: 1 MRANIPVRNFQEFYDAFGVKKGDSMYLKPEKRLTLW 36
+RANIP N +EFY+ F VK+ D MY PEKRL +W
25 Sbjct: 592 LRANIPPTNLEEFYETFDVKETDQMYRAPEKRLKIW 627

There is also some homology to SEQ ID 2384:

- Identities = 13/36 (36%), Positives = 25/36 (69%)

30 Query: 1 MRANIPVRNFQEFYDAFGVKKGDSMYLKPEKRLTLW 36
+R N+ + NF F++ F +K+GD+M+ P+ R+ +W
Sbjct: 596 LRTNVILTINFDAFHETFDIKEGDAMWRAPKDRVIIW 631

- Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1238

A DNA sequence (GBSx1314) was identified in *S.galactiae* <SEQ ID 3833> which encodes the amino acid sequence <SEQ ID 3834>. This protein is predicted to be endopeptidase O. Analysis of this protein sequence reveals the following:

- 40 Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45 bacterial cytoplasm --- Certainty=0.3801(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 50 >GP:AAA16168 GB:L18760 endopeptidase [Lactococcus lactis]
Identities = 118/268 (44%), Positives = 174/268 (64%), Gaps = 6/268 (2%)

-1385-

Query: 1 MGDYYGKKYFGEAAKDVHMAKKIINVYKTRLKNNTWLSNTKAMAICKLDNMRLMIGY 60
 +G +Y GKKYFGEAAK DV+ M +I VY+ RL N WLS+ T AI+KLD + IG+
 Sbjct: 321 IGLFYGKKYFGEAAKADVKRMVTAMIKVYQVRLSKNEWLSQETAKEAIEKLDAITPFIFG 380

Query: 61 PEDYPDLYRQYQFDSKASFFENNDNYRKLSNKKTFEEFNQSNQREHWQMSANAVNAYNDP 120
 P+ P++Y + + S S +E+ + K+ +TFE+F++ + W M A+ VNAY P
 Sbjct: 381 PDKLPEIYSRLKTTTS-GSLYEDALKFKDKILTARTFEKFSQEDVDKTSWHMPAHMVNAYYSP 439

Query: 121 NTNSIVFPAAIFQSPLYDKTKTVSQNYGAIGAIIGHEISHSFDINGMKYDEKGNLHDWWT 180
 ++N+IVFPAAI Q+P Y ++ SQNYG IGA+I HEISH+FD NG ++D++GNL+ WW
 Sbjct: 440 DSNTIVFPAAILQAPFYSLQSSSQNYGGIGAVIAHEISHAFDNGAQFDKEGNLKNWWL 499

Query: 181 KEDLKHKKKTQAMIDQWDGLKADGGKVDGKLTAEINADNGGVMASLEALKTEKIQTIK 240
 ED + +++K + MI +DG++ + G +GKL ++ENIAD GG+ A+L A K EK +K
 Sbjct: 500 DEDYEAFFBKQKEMIALFDGVETEAGFANGKLIVSENADQGGITAALTAAKDEKDVLDK 559

Query: 241 NFLNHGQVFGVKKQPKNKVSPQFSQMF 268
 F + K + K S +F QM +
 Sbjct: 560 AFFSQW-----AKIWRMKASKEFQQLL 582

There is also homology to SEQ ID 2384:

Identities = 110/253 (43%), Positives = 161/253 (63%), Gaps = 1/253 (0%)

Query: 1 MGDYYGKKYFGEAAKDVHMAKKIINVYKTRLKNNTWLSNTKAMAICKLDNMRLMIGY 60
 +G +Y + F AK DVE ++I VYK+RL+ WL+ T+ AI KL+ + IGY
 Sbjct: 324 LGLWYAGQKFSPEAKADVESKVARMIEVYKSRLETADWLAPATREKAITKLNVTIPHIGY 383

Query: 61 PEDYPDLYRQYQFDSKASFFENNDNYRKLSNKKTFEEFNQSNQREHWQMSANAVNAYNDP 120
 PE P+ Y + D S EN N K++ T+ ++N+ R W M A+ VNAY D
 Sbjct: 384 PEKLPETYAKKVIDESLSLVENAQNLAKTIAHTWSKWNKPVDRSEWHMPAHLVNAYYDL 443

Query: 121 NTNSIVFPAAIFQSPLYDKTKTVSQNYGAIGAIIGHEISHSFDINGMKYDEKGNLHDWWT 180
 N IVFPAAI Q P Y ++ S NYG IGA+I HEISH+FD NG +DE G+L+DWWT
 Sbjct: 444 QQNQIVFPAAILQEPFYSLDQSSSANYGGIGAVIAHEISHAFDTNGASFDEHGSNDWWT 503

Query: 181 KEDLKHKKKTQAMIDQWDGLKADGGKVDGKLTAEINADNGGVMASLEALKTEKIQTIK 240
 +ED +K++T ++ Q+DGL++ G KV+GKLT++EN+AD GGV +LEA ++E+ + +
 Sbjct: 504 QEDYAAFKERTDKIVAQFDGLESYGAKVNGKLTSENVDLGGVACALEAAQSEEDFSAR 563

Query: 241 N-FLNHGQVFGVK 252
 + F+N ++ +K
 Sbjct: 564 DFFINEFATIWRMK 576

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1239

A DNA sequence (GBSx1315) was identified in *S. agalactiae* <SEQ ID 3835> which encodes the amino acid sequence <SEQ ID 3836>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9691> which encodes amino acid sequence <SEQ ID 9692> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1386-

>GP:AAC35997 GB:AF019410 endopeptidase O [Lactobacillus helveticus]
 Identities = 85/315 (26%), Positives = 146/315 (45%), Gaps = 8/315 (2%)

5 Query: 46 NVSPRENLYRAVNDNWLANTKLKQGQTSVNSFSEIEDKLKQLLVSDMAKMASGKIETTN- 104
 N P++NLY AVN WL+ ++ QTS +E++ K+++ ++ D A +ASGK + +
 Sbjct: 20 NAKPQDNLYLAVNSEWLSKAEIPADQTSAGVNTELDIKIEKRMMKDFADIASGKEKMPDI 79

10 Query: 105 DEQKKMVAYYKQGMDFKTRDKNGLKPLKPVLLQKLEAVSSMKDFQSLAHDFVMSGFVLPFG 164
 + K +A YK +F RD P++ LQK+ + + F+ A + M + LPP
 Sbjct: 80 RDFDKAIALYKIAKNFDRDAEKANPIQNDLQKILDILNFDKFKDNATELFMGYPYALPFV 139

15 Query: 165 LTVETNARDNSQQLVLRQAPALLESPDQYKKGKGEAKLSAYRTSAMALLKQAGKSNI 224
 V+ + ++ L L YK E + L ++ LL+ AG
 Sbjct: 140 FDVDADMKNITDFNVLFHFGGPSTFLPDITTYK--TPEAKKLLDILEKQSINLLEMAGIGKE 197

20 Query: 225 EDRKLVKQAIADFRLLESEKTSQVDSKITAESETAAGRYNPESMETVHNYAKEFDKELIE 284
 E R V+ A+AFD+ LS+ K T E A YNP S+ K FD + ++
 Sbjct: 198 EARVYVQNALAFDQKLSKV-----KSTEWSDYAAIYNPVSLEFLAKFKSFDMAFLK 252

25 Query: 285 KLVGPTNKAVNVEDKTYFKQVNDVINSKQLANMKAMMISMLVDQSDFLGEQNROAASAF 344
 ++ + V V + + +++IN +K WM++ + + +L + R AA F
 Sbjct: 253 TILPEKVERVIVMEPRFLDHADELINPANFDEIKGWMLVKYINSVAKYLSQDFRAAAPPF 312

Query: 345 KNVASGLTQIESKEK 359
 SG ++ S+ K
 Sbjct: 313 NQAISGTPELPSQIK 327

A related GBS gene <SEQ ID 8763> and protein <SEQ ID 8764> were also identified. Analysis of this protein sequence reveals the following:

30 Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: 5.41
 GvH: Signal Score (-7.5): -1.39
 Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.

35 ALOM program count: 0 value: 2.76 threshold: 0.0
 PERIPHERAL Likelihood = 2.76 151
 modified ALOM score: -1.05

40 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

SEQ ID 8764 (GBS12) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 7; MW 65kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 3; MW 39kDa).

The GST-fusion protein was purified as shown in Figure 189, lane 4.

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1240

A DNA sequence (GBSx1317) was identified in *S.galactiae* <SEQ ID 3839> which encodes the amino acid sequence <SEQ ID 3840>. Analysis of this protein sequence reveals the following:

55 Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.75 Transmembrane 301 - 317 (299 - 317)

-1387-

----- Final Results -----

bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB42180 GB:A67181 unnamed protein product [unidentified]

Identities = 245/771 (31%), Positives = 410/771 (52%), Gaps = 80/771 (10%)

10 Query: 22 VRVIVEFNKESILDYATEQKKIVAQLNQADVEKKLQSIKQEQDKVLKNIKESVHFDSSKV 81
 VRVIV NK + D+ ++ + A + + +E+ +K Q+KV+K +E+ +KV
 Sbjct: 97 VRVIVSLNKSAAFDHTSKPTGSAASVKK--IEQASDQVKDQGEKVIKQVEE---ITGNKV 151

15 Query: 82 KR-YDAIINGVALDIQAQEIEKLKTIADVRRVYVSQEYVQTKPLSSSGQLIGLPEVWNN 140
 +R + ++N ++D+ +I+K+K + V+ V + Y P S+ Q+ + +VW
 Sbjct: 152 RRQFGYLVNAFSDIDMLDDIDKVKDLPQKKNVTPVKVY---HPTDESADQMAQVDVWQE 208

20 Query: 141 SQYKGEFTVVAVIDSGVDFKHQALKIKEPNRAKYNKTSIE----KLIHEKNLKGKFYSEK 196
 + KGEV V+++ID+G+D HQ LK+ +K+ +E KL H GK+Y+EK
 Sbjct: 209 QKLKGEGMVISIIDTGIDSSHQDLKLDGSGVSTALSKSEVESDKSKLGH-----GKYYTEK 263

25 Query: 197 VPGYNYDYNDNLKDS-YGVMHGMHVTGIVGANDDNQKLYGVAPNAQILAMKVFSDDDQ 255
 VPGYNY D ND + D+ G MHG HV GI GAN ++ GVAP+AQ+LAMKVFS++ +
 Sbjct: 264 VPGYNYADKNDQIVDNGCGEMHGQHVAGIAGANG---QVKGVAPDAQLLAMKVFSNNAK 320

30 Query: 256 NPTFTFDVWLKALDDAILLKADVVNMSLGTAPAGFVHEGKDYPELEVIARACKAGIVIAVA 315
 N + D + A++D++ L ADV+NMSLG+ + V G P+ + +A+A +AG++ ++
 Sbjct: 321 NSGAYDDDIISAIEDSVKLGADVINMSLGSVSSDV--GPSDPQQQAVAKASEAGVINVIS 378

35 Query: 316 AGNE---GNITDNTYGVKPLAENYDTALIANPALDDNTLAVASMENLKKHAHVLFK-- 370
 AGN G+ DGN +E + + P + + L VAS EN K +K +
 Sbjct: 379 AGNSGVAGSTADGNPVMNTGTSE---LSTVGTGPTPDALTVAASAENSKVTTDTVKDELG 435

40 Query: 371 -----DKKSGTEVTEVINLHVAPNASKTIIIGLAVDLGAGAPSELS--KHFDLSGKIA 420
 + K +VT + + + K + VD+G G + + K ++ G++A
 Sbjct: 436 GVTFFSSNSELKGAQVTTQLESNYSVLTKKLK---VDMGLGGADDTAEKKAIEVKGQLA 492

45 Query: 421 MLEIPEDNKSNGFLEKVQAITKLNPAAILLYNNAKVDDLGSQLLVESEAAKFNIARITR 480
 +++ + F KV A I++YN+ D L S L + +++
 Sbjct: 493 VVK----RGAYTFSKAVANAKAAGAAGIVTYNSE--DDGLLSMSLDDKTFTPLGMSKADG 546

50 Query: 481 STY----NNIKNNSNKIITILITERQAIDNSLAGQLSSYSSWGPTPDLRLKPEITAPGGHI 536
 + ++ + K T L IDNS AG++S ++SWGPTP+L KPEITAPGG I
 Sbjct: 547 KFWLKQKKVSRSLKFGTAL-----IDNSRAGKMSDFTSWGPTPELDFKPEITAPGGKI 601

55 Query: 537 FSTVEDNQYADKSGTSMAPQVAGAAVLKQYITDKKIPV--DNAADFIKLLLMNTAQPI 594
 +S DN+Y SGTSM+P VAG+ A++ Q I + + + + F K MNT+ P+
 Sbjct: 602 YSLANDNKYQQMSGTSMASPFVAGSEALILQGIKKQGLNLSGEELVQFAKNSAMNTSHPV 661

60 Query: 595 IN-KQSKDGKTPYFVRQOGSGAMNLAALVTTTVATVIGTNDNNADGKLELREL-KEKKF 652
 + + +K+ +P R+QSGG +N+ A+ TV N +G L+E+ ++ F
 Sbjct: 662 YDTEHTKEIISP---RRQSGEINVKDAINNTVEVKAA-----NGNGAALKEIGRQTTF 713

65 Query: 653 KARILLRNFGKTNKTYIISSEA--IADPVDEKGFRTQNSEHLVSKKADAVTRKVTVEAGK 710
 K + L N GK +TY + + + K +++ +V + T KVTV+ G+
 Sbjct: 714 K--VTLTNHGKKAQTYAVDNYGGPYTQATEAKSGEIYDTK-IVKGQLTTETPKVTVQPG 770

Query: 711 TLAVDLDVDYSDAEALTRNNFLEGYLNK-DTEGVADLHLPFLGFGYSWTE 760
 +VD+ + + R NF+EGY+ + + +L LP++GF+GS+++
 Sbjct: 771 --SVDVSFTLTLPSYFQRQNFVEGYVGFVAKDQATPNLVLPYMGFFGSYSQ 819

A related GBS gene <SEQ ID 8767> and protein <SEQ ID 8768> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -8.37

65

GvH: Signal Score (-7.5): -6.06

>>> Seems to have no N-terminal signal sequence

ALOM program count: 1 value: -1.75 threshold: 0.0

*** Reasoning Step: 3

----- Final Results -----

15

ORF00677(358 - 3159 of 3255)

%Match = 15.5

%Identity = 33.3 %Similarity = 54.6

Matches = 275 Mismatches = 343 Conservative Sub.s = 176

KAVTVTKPQGAVA EKATPAVPKPKQVRVIVEFNKESILDYATEQKKTV AQLNQADVEKKLQS I KQE QDKVLKNIEKS VHF
| | | | : | | : : | : : : | : : : | : | : | :
SKFQEA AKEQRQASGQAVSKKNES SVRVIVSLNKSAAFDH TSKPTGS AASV--KKIEQ ASDQVKD GQEKVIKQVEE---I

90 100 110 120 130 140

[illegible]

GVDFKHQALKIKEPNRAKYNTKSIEKLIHEKNLKGKFYSKVPGYNYDYDNDLKDS-YGVMHGMHVTGIVGANDDNQK
|:| ||||: : |: | | ||:||||| | | : | || || || |
GIDSSHQDLKLDSGVSTALSKEVESDKS-KLGHGKYTEKVPFYGYNYADKNDQIVDNGCGEMHGQHVAGIAGAN---GQ

795 825 855 885 915 942 972 1002

240 250 260 270 280 290

1032 1062 1092 1122 1152 1182 1212 1242
LYGVAPNAQILAMKVFSDDQQNPTTFDVLWKALDDAILLKADVNMSLGTPAGFVHEGKDYPELEVIARACKAGIVIAV
: ||||:|:|||||:::| :| :|::||:| |||:||||: :| | |: :|:| :||:: :
VKGVPADQQLLAMKVFSNNAKSGAYDDDIISAIEDSVKLGADVNMSLGSSSDV--GPSDPQQQAVAKASEAGVINVI
310 320 330 340 350 360 370

```

1272      1302      1326      1356      1386      1416      1656
AAGNEGNI TDGNTYGVKPLAENYD TAL-- IANPALDDNTLAVASMENLKKH A HVLKFKDKKSGTEVTEV~~~AAILLYN
: ||| |   : |   :   :   :   :   :   : ||| |||
SAGNSG--VAGSTADG N P V N N T G T S E L S T V G T P G V T P D A L I V A S A E N S K-----
              390      400      410      420

```

1686 1716 1746 1776 1806
 NAKVKDDLGSOLLVESEAAKFNIARITRSTYNNIKNNSNKIITILTERQA-----
 | :: : :| |:::| : :| | :
 -----VTTDTVKDELGGVTFSSNSSELKG-AAQVTTQLESNYSVLTKKLKLVDMGLGGADDDYT~~~FWLKQQ

-----IDNSLAGQLSSYSSWGPTDLRLKPEITAPGGHIFSTVEDNQYADKSGTSMAAPOVAGAAAVLKQY
 ||| |::| ::|||::| :|||::| :| ||: |||::| |||::| ::| :
 KKVRASRLKFGTALIDNSRAGKMSDFTSWGPTELPDLFKPEITAPGGKIYSLANDNKYQQMSGTSMASPFVAGSEALIQG

570 580 590 600 610 620 630

65

-1390-

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA95000 GB:AB042239 Paa [Streptococcus criceti]
Identities = 55/166 (33%), Positives = 81/166 (48%), Gaps = 24/166 (14%)

5   Query: 5   KKTDKFGFRKSKVCRSLCGALLGTVAVVSLATASTEIHADEATTSPFTTVTKVPQPVQADT 64
      K+ + FGFRKSK+ +SLCGALLGT VVS+ A A++ TTS T+ DT
      Sbjct: 2   KRKETFGFRKSKISKSLCGALLGTAIVVSV--AGQRAEAEDMTTSTTSA-----VDT 51

10  Query: 65   TALNTSKTHSTQATTTTPEAKENKVKSEITVQSESrv--MPRD-KVVERPETVKASVNS- 120
      TA+ ++T + +A + ++ Q+E + MP D E E VK++ +
      Sbjct: 52   TAVVGTETGNPATNLPEKQADSSSQAEASQAQAEQKTGSMFPVDVATTELDEAVKSAEAG 111

      Query: 121 -DVSQPIITTPPTI-----NEKTVEIPNLAQDTKKVAPKVTVTPE 159
      VSQ T T+ +EK+ EI D K A + +T E
15  Sbjct: 112   VTVSQDETVDKGTGVTGSQEADEKSGEI---KADYSKQAEITIKITTE 154

```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3842 (GBS222) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 6; MW 22kDa).

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1242

- 25 A DNA sequence (GBSx1319) was identified in *S.agalactiae* <SEQ ID 3843> which encodes the amino acid sequence <SEQ ID 3844>. This protein is predicted to be CylK. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
30      bacterial cytoplasm --- Certainty=0.3738(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1243

- 40 A DNA sequence (GBSx1320) was identified in *S.agalactiae* <SEQ ID 3845> which encodes the amino acid sequence <SEQ ID 3846>. This protein is predicted to be CylJ. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45      bacterial cytoplasm --- Certainty=0.1143(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

- 50 A related GBS nucleic acid sequence <SEQ ID 9689> which encodes amino acid sequence <SEQ ID 9690> was also identified.

-1391-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1244

- 5 A DNA sequence (GBSx1321) was identified in *S.agalactiae* <SEQ ID 3847> which encodes the amino acid sequence <SEQ ID 3848>. Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0913(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1245

- 20 A DNA sequence (GBSx1322) was identified in *S.agalactiae* <SEQ ID 3849> which encodes the amino acid sequence <SEQ ID 3850>. This protein is predicted to be CylI (fabF). Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
25 INTEGRAL Likelihood = -2.39 Transmembrane 721 - 737 (721 - 738)
INTEGRAL Likelihood = -1.97 Transmembrane 326 - 342 (326 - 343)
INTEGRAL Likelihood = -0.43 Transmembrane 534 - 550 (534 - 550)

- 30 ----- Final Results -----
bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 35 A related GBS nucleic acid sequence <SEQ ID 9687> which encodes amino acid sequence <SEQ ID 9688> was also identified.

There is also homology to SEQ ID 3852.

A related GBS gene <SEQ ID 8769> and protein <SEQ ID 8770> were also identified. Analysis of this protein sequence reveals the following:

- 40 Lipop: Possible site: -1 Crend: 8
McG: Discrim Score: 1.08
GvH: Signal Score (-7.5): -5.97
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
ALOM program count: 3 value: -2.39 threshold: 0.0
45 INTEGRAL Likelihood = -2.39 Transmembrane 712 - 728 (712 - 729)
INTEGRAL Likelihood = -1.97 Transmembrane 317 - 333 (317 - 334)
PERIPHERAL Likelihood = 3.45 492
modified ALOM score: 0.98

-1392-

*** Reasoning Step: 3

----- Final Results -----

5 bacterial membrane --- Certainty=0.1956 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

SEQ ID 8770 (GBS361) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 4; MW 84kDa).

10 GBS361-His was purified as shown in Figure 213, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1246

15 A DNA sequence (GBSx1323) was identified in *S.agalactiae* <SEQ ID 3853> which encodes the amino acid sequence <SEQ ID 3854>. This protein is predicted to be CylF. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

20 bacterial cytoplasm --- Certainty=0.3766 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1247

30 A DNA sequence (GBSx1324) was identified in *S.agalactiae* <SEQ ID 3855> which encodes the amino acid sequence <SEQ ID 3856>. This protein is predicted to be CylE. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

35 bacterial cytoplasm --- Certainty=0.3498 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1248

45 A DNA sequence (GBSx1325) was identified in *S.agalactiae* <SEQ ID 3857> which encodes the amino acid sequence <SEQ ID 3858>. This protein is predicted to be ABC transporter homolog CylB. Analysis of this protein sequence reveals the following:

-1393-

Possible site: 56

>>> Seems to have an uncleavable N-term signal seq

5 INTEGRAL Likelihood = -13.90 Transmembrane 271 - 287 (263 - 291)
 INTEGRAL Likelihood = -10.30 Transmembrane 17 - 33 (14 - 43)
 INTEGRAL Likelihood = -8.60 Transmembrane 114 - 130 (106 - 138)
 INTEGRAL Likelihood = -6.69 Transmembrane 152 - 168 (149 - 178)
 INTEGRAL Likelihood = -1.97 Transmembrane 186 - 202 (185 - 202)

----- Final Results -----

10 bacterial membrane --- Certainty=0.6562(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 A related GBS nucleic acid sequence <SEQ ID 9685> which encodes amino acid sequence <SEQ ID 9686> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1249

20 A DNA sequence (GBSx1326) was identified in *S.galactiae* <SEQ ID 3859> which encodes the amino acid sequence <SEQ ID 3860>. This protein is predicted to be ABC transporter homolog CylA. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

25

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.4122(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

A related GBS nucleic acid sequence <SEQ ID 9683> which encodes amino acid sequence <SEQ ID 9684> was also identified. A further related GBS gene <SEQ ID 8771> and protein <SEQ ID 8772> were also identified. Analysis of this protein sequence reveals homology to membrane protein ABC transporters.

35 A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9085> which encodes the amino acid sequence <SEQ ID 9086>. An alignment of the GAS and GBS sequences follows:

Score = 85.4 bits (208), Expect = 1e-18

Identities = 68/271 (25%), Positives = 129/271 (47%), Gaps = 17/271 (6%)

40 Query: 39 KGFTEQHVLKDINFDVYKGDFFGIVGRNGSGKSTLLKIIISQIYVPEKGQVT--VDGKMVS 96
 K + L+DIN +G F+G++G NG+GK+TL ++ Q + G + VDGK +S
 Sbjct: 10 KKYGSFEALRDINLIFEEGKFYGLLGPNAGAKTTLFNLLIQNFQKQTSGDIKWEVDGKPLS 69

 Query: 97 -----FIELGVGF-----NPELTGRENVMNGAMLGFTKDEVDDMYNDIVDFAELHHFMNQ 147
 + +G+ F + LT EN+ GA+ G +K +V + D+ + ++ Q
 45 Sbjct: 70 IKDFYRHIGIVFQSNRLDDNLTVEENLISRCALYGLSKSQVRNRLKDLQTYLDITAIIKKQ 129

 Query: 148 KLKNYSSGMQVRLAFSVAIKAQGDVLLILDEVLAVGDEAFQRKCNDYFME-RKDSGKTTIL 206
 K + S G + ++ + A+ Q +L+LDE D +R D + + S T +L
 50 Sbjct: 130 KYGSLSGGQKRKVDIARALLPQPSLLLLDEPTTGLDPQSRDLWDAIAQLNQSQMTVVL 189

 Query: 207 VTHDMGAVKICYCNRAVLIEDGLVKAYGEPFDVANQYSVDNTETA-EDAMNAEKISVSDIA 265
 +TH + + C+ ++ +G + G+ Q+S N + + ++S++D
 Sbjct: 190 ITHYLEEMSA-CDVLNVLIEGNIYYSGDIKSFIEQHSTNLNVVLKPEKSLDQLSIADFV 248

55 Query: 266 KDLKVSILISNPRITFNDTITFEVSYEVLKDD 296

-1394-

K ++S I D I+ E +V+ D+
 Subject: 249 N--KCQVLSEREIVFKD-ISVEEMMQVISDN 276

There is also homology to SEQ IDs 358, 482, 644, 686, 1832, 2529, 2720, 3882, 4028, 4104, 4280, 5090,
 5 5498, 6034, 6500.

SEQ ID 8772 (GBS83) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 2; MW 37.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 5; MW 62.6kDa) and in Figure 28 (lane 3; MW 62.6kDa).

10 GBS83-GST was purified as shown in Figure 195, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1250

A DNA sequence (GBSx1327) was identified in *S.agalactiae* <SEQ ID 3861> which encodes the amino
 15 acid sequence <SEQ ID 3862>. This protein is predicted to be acyl carrier protein homolog AcpC. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3451(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1251

A DNA sequence (GBSx1328) was identified in *S.agalactiae* <SEQ ID 3863> which encodes the amino
 30 acid sequence <SEQ ID 3864>. This protein is predicted to be CylG (fabG). Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2651(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 There is also homology to SEQ ID 3866.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1395-

Example 1252

A DNA sequence (GBSx1329) was identified in *S.agalactiae* <SEQ ID 3867> which encodes the amino acid sequence <SEQ ID 3868>. This protein is predicted to be CylD. Analysis of this protein sequence reveals the following:

5 Possible site: 60
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.2030(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1253

A DNA sequence (GBSx1330) was identified in *S.agalactiae* <SEQ ID 3869> which encodes the amino acid sequence <SEQ ID 3870>. Analysis of this protein sequence reveals the following:

20 Possible site: 14
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.3219(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1254

A DNA sequence (GBSx1331) was identified in *S.agalactiae* <SEQ ID 3871> which encodes the amino acid sequence <SEQ ID 3872>. Analysis of this protein sequence reveals the following:

35 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.97	Transmembrane	231 - 247 (226 - 251)
INTEGRAL	Likelihood = -7.06	Transmembrane	141 - 157 (134 - 164)
INTEGRAL	Likelihood = -2.76	Transmembrane	28 - 44 (26 - 44)
INTEGRAL	Likelihood = -1.38	Transmembrane	123 - 139 (121 - 139)
40 INTEGRAL	Likelihood = -0.32	Transmembrane	199 - 215 (199 - 215)

 ----- Final Results -----
 45 bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB88836 GB:AL353832 putative integral membrane transport
 protein. [Streptomyces coelicolor A3(2)]
 50 Identities = 68/264 (25%), Positives = 123/264 (45%), Gaps = 10/264 (3%)

-1396-

Query: 6 RMHFIFIKQYMKQIMEYKIDFFVGVGLGVFLTQGLNLLFLNVLFQHIPLSLEGWTFQQIAFI 65
 R + + + + M Y+ F + G F L+ + + ++F + +L G++ ++AF+
 Sbjct: 34 RAYGLIAGMWIRSTMAYRTSFALTAFGNFAMTALDFVAILLMFSRVDALGGYSLPEVAFL 93

5

Query: 66 YGFSLLPKGIDHLEFFDNLWALGQRLIRKGEFDKYLTRPISPLFHVIVETVQVDALGELLV 125
 YG S + G+ L ++ LG+R +R G D L RP L V + F + LG ++
 Sbjct: 94 YGLSGVSFGLADLAIGSMERLGRR-VRDGTLDTLVLRPAPVLAQVAADRFALRRGRVVQ 152

10

Query: 126 GFILL--STTVSSISWTVPKVLLFIFIIPFATLIYTSKTIATSSIAFWTKQSGAVIYIF- 182
 G ++L + V I WT KVLL + I+ ++ +A + F + + V F
 Sbjct: 153 GLLVLGYALVVVDIDWTAAKVLLLPVALISGAGIFCAVFVAAGAFQFAAQDASEVANAF 212

15

Query: 183 YMFNDFAKYPVAIYNLLRWIISFVIPFAFTAYYPAAYFLQDRNVYFNIGGVI-----LI 237
 Y +YP ++ L +FV+P AF + PA+Y L R ++ G + L
 Sbjct: 213 YGGTTLQYPPVTFALDLVRGATFVLPLAFVNWLPASYVL-GRPYPLDLPGWVAFTPPLA 271

Query: 238 SLISFMVSLILWHKGVVEYESAGS 261
 + ++ + W G+ Y S GS
 Sbjct: 272 AAACCALAGLAWRAGLRSYRSTGS 295

20

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3873> which encodes the amino acid sequence <SEQ ID 3874>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.86	Transmembrane	227 - 243 (225 - 251)
INTEGRAL	Likelihood = -7.22	Transmembrane	141 - 157 (133 - 164)
INTEGRAL	Likelihood = -6.37	Transmembrane	123 - 139 (114 - 140)
INTEGRAL	Likelihood = -2.97	Transmembrane	26 - 42 (26 - 49)

30

----- Final Results -----

bacterial membrane	---	Certainty=0.4545(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

35

The protein has homology with the following sequences in the databases:

>GP:CAB88836 GB:AL353832 putative integral membrane transport protein. [Streptomyces coelicolor A3(2)]
 Identities = 69/262 (26%), Positives = 125/262 (47%), Gaps = 10/262 (3%)

40

Query: 8 HAIFIKQYLKQIMEYKVDFFVGVGLGVFLTQGLNLLFLSVLFQHIPLSLEGWTFEQIAFIY 67
 + + + + M Y+ F + G F L+ + + ++F + +L G++ ++AF+YG
 Sbjct: 36 YGLIAGMWIRSTMAYRTSFALTAFGNFAMTALDFVAILLMFSRVDALGGYSLPEVAFLYG 95

45

Query: 68 FCLIPKGIDHLEFFDNLWALGQRLVRKGEFDKYLTRPISPLFHVIVETVQVDALGELLVGV 127
 + G+ L ++ LG+R VR G D L RP L V + F + LG ++
 Sbjct: 96 LSGVSFGLADLAIGSMERLGRR-VRDGTLDTLVLRPAPVLAQVAADRFALRRGRVVQGL 154

50

Query: 128 ILL--VTTAGSIVWTLPKVLLFILVIPFATLIYTSKTIATASISFWTKQSGAVIYIF-YM 184
 ++L I WT KVLL + + I+ ++ +A + F + + V F Y
 Sbjct: 155 LVLGYALVVVDIDWTAAKVLLLPVALISGAGIFCAVFVAAGAFQFAAQDASEVANAFY 214

55

Query: 185 FNDFSKYPMSTYHSFLRWLISFIIPFAFTAYYPAAYFLTGQHLLEFNIGGLV-----VVSL 239
 +YP +++ L +F++P AF + PASY L G+ ++ G V + +
 Sbjct: 215 GTTMLQYPPVTFALDLVRGATFVLPLAFVNWLPASYVL-GRPYPLDLPGWVAFTPPLAA 273

Query: 240 LVLALSLKLWKWGLDAYESAGS 261
 AL+ W+ GL +Y S GS
 Sbjct: 274 ACCALAGLAWRAGLRSYRSTGS 295

60

An alignment of the GAS and GBS proteins is shown below.

Identities = 208/261 (79%), Positives = 238/261 (90%)

Query: 1 MTKYQRMHFIFIKQYMKQIMEYKIDFFVGVGLGVFLTQGLNLLFLNVLFQHIPLSLEGWTFQ 60

-1397-

M K + MH IFIKQY+KQIMEYK+DF VGVLG VFLTQGLNLLFL+VLFQHIPSLEGWTF+
 Sbjct: 1 MAKLRCHAIPIKQYLKQIMEYKVDFFVGVLG VFLTQGLNLLFLSVLFQHIPSLEGWTFE 60
 Query: 61 QIAFIYGFSLLEPKGIDHLFFDNLWALGQRLIRKGEFDKYLTRPISPLFHVLVETTFQVDAL 120
 5 QIAFIYGF L+PKGIDHLFFDNLWALGQRL+RKGEFDKYLTRPISPLFHVLVETTFQVDAL
 Sbjct: 61 QIAFIYGFCLIPKGIDHLFFDNLWALGQRLVRKGEFDKYLTRPISPLFHVLVETTFQVDAL 120
 Query: 121 GELLVGFILLSTTVSSISWTVPKVLLFIFIIIPFATLIYTSLKIIATSSIAFWTKQSGAVIY 180
 GELLVG ILL TT SI WT+PKVLLFI +IPFATLIYTSLKIIAT+SI+FWTKQSGAVIY
 10 Sbjct: 121 GELLVGILLVTTAGSIVWTLTPKVLLFILVIPFATLIYTSLKIIATASISFWTKQSGAVIY 180
 Query: 181 IFYMFNDFAKYPVAIYNLLRWIISFVIPLEAFTAYYPAAYFLQDRNVYFNIGGVILISLI 240
 IFYMFNDF+KYP++IY++ LRW+ISF+IPFAFTAYYPA+YFL +++ FNIGG++++SL+
 15 Sbjct: 181 IFYMFNDFSKYPMSTYHSLRWLISFIIPFAFTAYYPAAYFLTGQHLLEFNIGGLVVVSL 240
 Query: 241 SFMVSLILWHKGVVYESAGS 261
 +SL LW G++ YESAGS
 Sbjct: 241 VLALSLKLWKWGLDAYESAGS 261

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1255

A DNA sequence (GBSx1332) was identified in *S. agalactiae* <SEQ ID 3875> which encodes the amino acid sequence <SEQ ID 3876>. Analysis of this protein sequence reveals the following:

25 Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -15.60 Transmembrane 147 - 163 (134 - 178)
 INTEGRAL Likelihood = -8.55 Transmembrane 119 - 135 (114 - 141)
 INTEGRAL Likelihood = -7.86 Transmembrane 238 - 254 (235 - 260)
 30 INTEGRAL Likelihood = -1.70 Transmembrane 215 - 231 (212 - 231)
 INTEGRAL Likelihood = -1.06 Transmembrane 61 - 77 (61 - 77)
 INTEGRAL Likelihood = -0.22 Transmembrane 27 - 43 (27 - 43)
 ----- Final Results -----
 35 bacterial membrane --- Certainty=0.7241(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB88837 GB:AL353832 putative integral membrane protein.
 [Streptomyces coelicolor A3(2)]
 Identities = 60/271 (22%), Positives = 118/271 (43%), Gaps = 13/271 (4%)
 Query: 6 RRYKPFISTGIQGLITYRVDFILYRIGDVIGAFVAFYLWKAVFDSSSQSLIQGFQLSDMI 65
 45 R Y + G + TYR + + + Y + A++D Q + G+ + +
 Sbjct: 7 RLYVAVAAGGFRRYATYRAATAAGVFTNTVFGLILVYTYLALWDEKQP--LGGYDQAQAV 64
 Query: 66 LYIIMS-FVTNLLTRTDSSFM--IGDEVKDGSIIMRLRPVHFAASYLFMEIGSRWLIFL 122
 ++ + + L F + + ++ G + + L RP +L ++G L
 50 Sbjct: 65 TFWVLGQALLAALAIGGGGFEDLMERIRITGDVAVDLYRPADLQLWWLAADVGRAVFQL 124
 Query: 123 SIGV-PFLLVITGVRLFLGTDLIQAIVLVVFIISIILAFLINFFFNICGFSAFVFKNL 181
 GV EF+ LF L + + + +++++LA ++ F SAF +
 55 Sbjct: 125 GRGVVPFVFG---SLFFPVALPREVSVWAAFLVAVVLAMVVGFAIRYLVALSAFWLLDG 180
 Query: 182 WGSNLLKNSLVAFMSGSLIPLTFFPKIVADILGFLPFSSLIYTPVMIIIGKYDGSQIVQA 241
 G + F SG L+PL FP ++ D++ LP+SSL+ P +++G+ D +
 Sbjct: 181 TGVITQMAWLAGLFCGMLLPLNVFPGVLGDVVRALPWSSLLQGPADVLLGEADP---LGT 237
 60 Query: 242 LLLQIFWLVIMVALSQLIWKVQLHITIQQG 272
 L Q W + ++AL +L+ + +QGG
 Sbjct: 238 YLFQASWAVALLALGRLVQSAATRRVVVQGG 268

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3877> which encodes the amino acid sequence <SEQ ID 3878>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 50
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -9.18    Transmembrane 252 - 268 ( 248 - 277)
      INTEGRAL    Likelihood = -7.22    Transmembrane 161 - 177 ( 151 - 187)
      INTEGRAL    Likelihood = -6.10    Transmembrane 133 - 149 ( 128 - 160)
10     INTEGRAL    Likelihood = -2.81    Transmembrane 213 - 229 ( 211 - 230)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

      >GP:AAF11144 GB:AE002002 conserved hypothetical protein [Deinococcus radiodurans]
      Identities = 56/268 (20%), Positives = 113/268 (41%), Gaps = 21/268 (7%)

20     Query: 15 MWSFWKRYRPFSLSAGIQELITYRVNFFLYRIGDVMGAFVAYYLWKAVFDSSKQSLINGFT 74
      M +FW++ R + + + YR ++ + + V +W S+ ING+T
      Sbjct: 1 MTNFWRKVRVLWAVSLASTLEYRAETIIWMLSGTLN-LVMMLVWMTQAKSAPGGQINGYT 59

      Query: 75 LSDMTFYIIMS FVTLLTKSDSSFMIGEEVKDGSIIIMRLLRPV-----HFAASYLFMEIG 129
      Y + +++ + L + + +++ G++ LL P+ FAA +
25     Sbjct: 60 PQAFAGYFLATWLVSQLLVVWVGWELDYKIRQGTLSPELLHPIDPLWREFAAH--LTDKA 117

      Query: 130 FRWIVLMSVGFPFLMVLVSGIKVMAGLSILQVLASSCLYLVSLLLAFL---INFYFNICFG 186
      FR P ++VL + + A L+ Q + Y L LA L + F + G
30     Sbjct: 118 FR-----LPIMLVL---LLIFAALTGAQFTSQWYPAVLGLALLGLCVRFLEWYTLG 167

      Query: 187 SSAFVFNKLVGWSNLLKNALVAFMSGSLIPLAFFPKMVSIVLSFLPFSSLVYTPVMIVIGK 246
      AF ++ + A G PL+F+P + + ++ PF ++ P ++ GK
      Sbjct: 168 LLAFWTESSSSFGVWLWLFYAAFPGMFAPLSFYPGWLQTLAAWTPFPYMLGLPAALLAGK 227
35     Query: 247 YSLSQIMVALSLQIFWLLVMVLSQVIW 274
      S ++ + + + WL VM ++ + +W
      Sbjct: 228 ASGAELRGAGVLLGWLAVMMLVRRWVW 255

```

40 An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 199/268 (74%), Positives = 236/268 (87%)

      Query: 5 WRRYKPFISTGIQGLITYRVDFILYRIGDVIGAFVAFYLVWKAVFDSSSQSLIQGFQSLSDM 64
      W+RY+PF+S GIQ LITYRV+F LYRIGDV+GAFVA+YLWKAVFDSS QSLI GF LSDM
45     Sbjct: 19 WKRYRPFSLSAGIQELITYRVNFFLYRIGDVMGAFVAYYLWKAVFDSSKQSLINGFTLSDM 78

      Query: 65 ILYIIMS FVTNLLTRTDSSFMIGDEVKDGSIIIMRLLRPVHFAASYLFMEIGSRWLIFLSI 124
      YIIMS FVT LLT++DSSFMIG+EVKDGSIIIMRLLRPVHFAASYLFMEIG RW++ +S+
50     Sbjct: 79 TFYIIMS FVTLLTKSDSSFMIGEEVKDGSIIIMRLLRPVHFAASYLFMEIGFRWIVLMSV 138

      Query: 125 GVPFLLVITGVRLFLGTDLIQAIVLVVFYIISIIAFLINFFFNICFGFSAFVFNKLVGWS 184
      G PFL+V++G+++ G ++Q + Y++S++LAFLINF+FNICFG SAFVFNKLVGWS
      Sbjct: 139 GFPFLMVLVSGIKVMAGLSILQVLASSCLYLVSLLLAFLINFFYFNICFGSSAFVFNKLVGWS 198

      Query: 185 NLLKNSLVAFMSGSLIPLTFPPKIVADILGFLPFSSLIYTFVMIIGKYDGSQIVQALLL 244
      NLLKN+LVAFMSGSLIPL FFPK+V+ +L FLPFSSL+YTFVMI+IGKY SQI+ AL L
55     Sbjct: 199 NLLKNALVAFMSGSLIPLAFFPKMVSIVLSFLPFSSLVYTFVMIIVIGKYSLSQIMVALSL 258

      Query: 245 QIFWLLVMVALSqliwkkvqlhitiqgg 272
      QIFWL+VMV LSQ+IWKKVQ H+TIQGG
60     Sbjct: 259 QIFWLLVMVLSQVIWKKVQYHLTIQGG 286

```

-1399-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1256

A DNA sequence (GBSx1333) was identified in *S.agalactiae* <SEQ ID 3879> which encodes the amino acid sequence <SEQ ID 3880>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2013(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9681> which encodes amino acid sequence <SEQ ID 9682> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF09790 GB:AE001882 ABC transporter, ATP-binding protein
[Deinococcus radiodurans]
Identities = 141/331 (42%), Positives = 201/331 (60%), Gaps = 34/331 (10%)

Query: 10 MIEVSHLQKNFIKTVKAPGLKGAFAQSFLRPEKHTFEAVKDLTFDVPKQIILGFIGANGAG 69
MIEV HL K+F + AV+D++F +P G+I+G++G NGAG
Sbjct: 46 MIEVRHLCKSFARK-----PAVQDISFSIPAGEIVGYLGPNGAG 84

Query: 70 KSTTIKMLTGILKPTSGFCRIDGKLPOENRONYVKDIGVVFQRTQLWWDLALQETTYTVL 129
KSTTIK+LTG+L P SG R+ G +P + R+ +V +G VFGQRT LWWDL ++E+ +L
Sbjct: 85 KSTTIKVLTLGLVPDSGEVRVGGGLVPWKQRRQHVARLGAVFGQRTTLWWDLPVRESLELL 144

Query: 130 KEIYDVPDKFEFRKRMFAFLNEVLELNDFIKDPVRTLSLGQMRADIAASLLHNPVKVLFDE 189
+ +Y VP F + +A E+LEL F+ P R LSLGQMRAD+AA+LLH+P++LFLDE
Sbjct: 145 RHVYRVPAARFAENLAGFTELELGPFLNTPARALSLGQMRADLAAALLHDPPELLFLDE 204

Query: 190 PTIGLDVSVKDNIRRAITQINQEEETILLTHDLSDIEQLCHRIFMIDRGQEIFDGTVS 249
PT+GLDV K+ IR + +N E T+LLTHDL D+E+L R+ MID G+ +FDG ++
Sbjct: 205 PTVGLDVVAKERIREFVKAVNAERGVTVLLTHDLGDVERLARRVMMIDTGRLLFDGPIA 264

Query: 250 QLKETFGKMKTL--SFDLRPGQEHISS-SLICKSEINIKRNDLVLDIQYDSSRYQTADII 306
+L+ +G + L F+ P Q + +L+G+ ++ Y S A I
Sbjct: 265 ELQARYGGERELWVEFEKAPAPALPGLTLLGQDGPRVR-----YGFSGAAAAPIA 315

Query: 307 QQTLDVSVRDLKMTDADIEDIIRRFYRNEL 337
Q T A VRDL + + ++E IRR Y L
Sbjct: 316 QVT-ALAPVRDLAVKEPEVEATIRRIYEGNL 345

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3881> which encodes the amino acid sequence <SEQ ID 3882>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3315(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 272/330 (82%), Positives = 305/330 (92%)

-1400-

Query: 8 MSMEVSHLQKNFIKTVKAPGLKGAFQSFRLPEKHTFEAVKDLTFDVPKGQILGFIGANG 67
 M MIEVSHLQKNF KT+K PGLKGA +SF+ P + FEAVKDL+F+VPKGQILGFIGANG
 Sbjct: 1 MVMIEVSHLQKNFSKTIKEPGLKGALKSFVHPPREIFEAVKDLSPFEVPGKQILGFIGANG 60
 Query: 68 AGKSTTIKMLTGILKPTSGFCRIDGKLPQENRQNYVKDIGVVFQRTQLWWDLALQETYT 127
 AGKSTTIKMLTGILKPTSG+CRI+GK+PQ+NRQ YV+DIG VFGQRTQLWWDLALQETY
 Sbjct: 61 AGKSTTIKMLTGILKPTSGYCRINGKIPQDNQYYVRDIGAVFGQRTQLWWDLALQETV 120
 Query: 128 VLKEIYDVPDKEFRKRMAFLNEVLELNDFIKDFVRTLSLGQMRADIAASLLHNPVKVLF 187
 VLKEIYDVP+K FRKRM FLNEVL+LN+FIKDFVRTLSLGQMRADIAASLLHNPVKVLF
 Sbjct: 121 VLKEIYDVPEKAFRKRMDFLNEVLDLNEFIKDFVRTLSLGQMRADIAASLLHNPVKVLF 180
 Query: 188 DEPTIGLDVSVKDNIRRAITQINQEEETTILLTTHDLSDIEQLCHRIFMIDRGQEIFDGT 247
 DEPTIGLDVSVKDNIRRAITQINQEEETTILLTTHDLSDIEQLC RI MID+GQEIFDGT
 Sbjct: 181 DEPTIGLDVSVKDNIRRAITQINQEEETTILLTTHDLSDIEQLCDRIIMIDKGQEIFDGT 240
 Query: 248 VSQKETFQKMKTLSTFDLRPGQEHSSSLIGKSEINIKRNDLVLDIQYDSSRYQTADIIQ 307
 V+QLK++FGKMK+LSF+L+PGQE + S +G +I ++R++L LDIQYDSSRYQTADIIQ
 Sbjct: 241 VTQLKQSFQKMKSLSFELKPGQEQQVVSQFMGLPDITVERHELSDIQYDSSRYQTADIIQ 300
 Query: 308 QTLADFSVRDLKMTDADIEDIIRRFYRNEL 337
 +T+ADF+VRD+KMTD DIEDI+RRFYR EL
 Sbjct: 301 KTMADFAVRDVKMTDVEDIEDIVRRFYRKEL 330

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1257

A DNA sequence (GBSx1334) was identified in *S. agalactiae* <SEQ ID 3883> which encodes the amino acid sequence <SEQ ID 3884>. This protein is predicted to be Fmt. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -9.39 Transmembrane 21 - 37 (8 - 39)
 INTEGRAL Likelihood = -7.75 Transmembrane 360 - 376 (359 - 381)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4758(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8775> which encodes amino acid sequence <SEQ ID 8776> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: 8.85
 GvH: Signal Score (-7.5): -3.75
 Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 2 value: -9.39 threshold: 0.0
 INTEGRAL Likelihood = -9.39 Transmembrane 21 - 37 (8 - 39)
 INTEGRAL Likelihood = -7.75 Transmembrane 353 - 369 (352 - 374)
 PERIPHERAL Likelihood = 4.24 92
 modified ALOM score: 2.38
 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4758(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1401-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA24012 GB:AB009635 Fmt [Staphylococcus aureus]
Identities = 72/279 (25%), Positives = 125/279 (43%), Gaps = 25/279 (8%)

5   Query: 49  LHRFMRKNNVNGMMIVSDNTGKPITISHGINRGEVETDIEN--NKLFPMASIQKLMGTGII 106
      + +++++ + NG + + +N GK + +S G      + E I+N N +F + S QK TG++
      Sbjct: 79  IDKYLQSSLFNGSVAIYEN-GK-LKMSKGYGYQDFEKGIKNTPTMTFLIGSAQKFSTGLL 136

10  Query: 107 IQRLIDQDVLSEDDRLSQFFPQVKGNSITIHQLLTHTSGLREKGVKVSPYLKNBREQLO 166
      +++L ++ ++ +D +S++ P K S I + L+ H SGL + K S KN + ++
      Sbjct: 137 LKQLEEEHKININDFVSKYLPWFKTSKPIPLKDLMLHQSGLYK--YKSSKDYKNLDQAVK 194

      Query: 167 FCLKHYNFVNK-KSWYYSNINFSFLTGIATQVTGRTYAEVLDDVIKNPLRLDDTQSYQSV 225
      K K Y++ N+ L + +VTG++YAE I +PL+L T Y
15  Sbjct: 195 AIQKRGIDPKYKKHMYNDGNVYLAKVIEEVTGKSYAENYYTKIGDPLKLQHTAFYD-- 252

      Query: 226 VNHDLVSPMRKNGKLNKINIF----NQVSTAYAGDFFTTPLNFWVLMRSFSKGYFFPT- 280
      + K N + N + YGAG+ + TP + L+ + F
20  Sbjct: 253 -EQPFKKYLAKGYAYNSTGLSFLRPNILDQYYGAGNLYMTPTDMGKLITQIQYKLFSPK 311

      Query: 281 -----DEYTKHQNDAISHYGGLYMHGRIVNSNGTFF 312
      + TK D Y G Y + NG FF
      Sbjct: 312 ITNPLLHEFGTKQYPD--EYRYGFYAKPTLNRLNGGFF 347
```

25 There is also homology to SEQ ID 3886.

A related GBS gene <SEQ ID 8773> and protein <SEQ ID 8774> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1 Crend: 6
McG: Discrim Score: 14.89
30  GvH: Signal Score (-7.5): -3.75
      Possible site: 25
      >>> Seems to have an uncleavable N-term signal seq
      ALOM program count: 1 value: -9.39 threshold: 0.0
      INTEGRAL Likelihood = -9.39 Transmembrane 14 - 30 ( 1 - 32)
35  PERIPHERAL Likelihood = 4.24 85
      modified ALOM score: 2.38

      *** Reasoning Step: 3

40  ----- Final Results -----
      bacterial membrane --- Certainty=0.4758(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

45 The protein has homology with the following sequences in the databases:

```
29.6/49.6% over 218aa
Bacillus cereus
GP|4127525| D-stereospecific peptide hydrolase Insert characterized

50  ORF00162(478 - 1083 of 1644)
GP|4127525|emb|CAA09676.1||AJ011526(67 - 285 of 389) D-stereospecific peptide hydrolase
{Bacillus cereus}
%Match = 5.8
%Identity = 29.5 %Similarity = 49.5
55  Matches = 62 Mismatches = 96 Conservative Sub.s = 42

      330      360      390      420      450      480      510      540
MILRRLFMVRKFLKSLLSFLIAVIATGISVACFFFIPEKNITPILLHRFMRKNNVNGMMIVSDNTGKPITISHGINR
      ::      : |      ::      :      :      |      ::      : ||      : ||
60  TCASLALLIAGSSLLYTTPTSIVKAEPTQNVSSSLQNTQDRDTSVKQAMRDTLQLGYPGILAKTSEGGKTIWGYAAGIAD
      20      30      40      50      60      70      80

      570      600      630      660      705      735      753
```

-1402-

```

5      GEVETDIENKLFPMASLQKLMGTGIIQRLIDQDVLSEDDRLSQFFPQV---KG--SNSITIHQLLTHTSGL---REKG
      : : : | : | : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
      LRTKKPMKTDFFRFRIGSVTKTFTTATVVLQLVGENRLKLDHIEDWLPQVIGNGYDGNKITIQEILNHTSGIAEYSRSKD
      100      110      120      130      140      150      160

10     807      834      864      894      924      954      978
      VKVSPYLKN--EREQLQFCLKHY-NFVNKKSWYYSNINFSFLTGIATQVTGRTYAELVDDVIKNPLRLDDT--QSYQSVV
      | : | : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
      VDFTDTKKSYSYAEELVKMGISFPDFAPGKGWSYSNTGYVLGLIEKVTGNSYAEVENRIIEPLELSNTFLPGNSSVI
      180      190      200      210      220      230      240

15     993      1023      1053      1083      1113      1143      1173      1203
      ---NH--DLVSPMRKNGKLNKINIFNQVSTAYGAGDFFTTPLNFWVLMRSFSKGYFFPTDEYTKHQNDALSHYYGGLYMH
      || | | : | : : | | | : | : : : : | : : : : : : : : : : : : : : : : : : : : : :
      PGTNHARGYVQP-DGASELKDVITYN-PSAGSSAGDMISTADDLNKFFSYLLGGKLLKEQQLKQMLTTVPTGKEGIDYG
      260      270      280      290      300      310      320

```

SEQ ID 8776 (GBS61) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 3; MW 68kDa).

GBS61-GST was purified as shown in Figure 195, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1258

A DNA sequence (GBSx1335) was identified in *S.agalactiae* <SEQ ID 3887> which encodes the amino acid sequence <SEQ ID 3888>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

```

```

30     ----- Final Results -----
           bacterial cytoplasm --- Certainty=0.2398(Affirmative) < succ>
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1259

A DNA sequence (GBSx1336) was identified in *S.agalactiae* <SEQ ID 3889> which encodes the amino acid sequence <SEQ ID 3890>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -5.57    Transmembrane    16 - 32 ( 13 - 33)

```

```

45     ----- Final Results -----
           bacterial membrane --- Certainty=0.3230(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1403-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1260

A DNA sequence (GBSx1337) was identified in *S.agalactiae* <SEQ ID 3891> which encodes the amino acid sequence <SEQ ID 3892>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3910(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1261

A DNA sequence (GBSx1338) was identified in *S.agalactiae* <SEQ ID 3893> which encodes the amino acid sequence <SEQ ID 3894>. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4239(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1262

A DNA sequence (GBSx1339) was identified in *S.agalactiae* <SEQ ID 3895> which encodes the amino acid sequence <SEQ ID 3896>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4349(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1404-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1263

A DNA sequence (GBSx1340) was identified in *S.agalactiae* <SEQ ID 3897> which encodes the amino acid sequence <SEQ ID 3898>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4962 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1264

A DNA sequence (GBSx1341) was identified in *S.agalactiae* <SEQ ID 3899> which encodes the amino acid sequence <SEQ ID 3900>. Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4014 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG38044 GB:AF295925 Orf28 [Streptococcus pneumoniae]
Identities = 23/35 (65%), Positives = 28/35 (79%)

Query: 9 LIHWEGNSGDKLIEHQTSATGWYQVDRSFSQPKG 43
L +WEGNSGDKL+E QT AT WYQ+++ FSQ G
Sbjct: 180 LTYWEGNSGDKLLERQTRATEWYQIEKGFSTNG 214

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1265

A DNA sequence (GBSx1342) was identified in *S.agalactiae* <SEQ ID 3901> which encodes the amino acid sequence <SEQ ID 3902>. Analysis of this protein sequence reveals the following:

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2036 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1266

A DNA sequence (GBSx1343) was identified in *S.agalactiae* <SEQ ID 3903> which encodes the amino acid sequence <SEQ ID 3904>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10933> which encodes amino acid sequence <SEQ ID 10934> was also identified.

SEQ ID 3904 (GBS153) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 25 (lane 3; MW 22kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 4; MW 47kDa).

GBS153-GST was purified as shown in Figure 198, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1267

A DNA sequence (GBSx1344) was identified in *S.agalactiae* <SEQ ID 3905> which encodes the amino acid sequence <SEQ ID 3906>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2036(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1268

A DNA sequence (GBSx1345) was identified in *S.agalactiae* <SEQ ID 3907> which encodes the amino acid sequence <SEQ ID 3908>. Analysis of this protein sequence reveals the following:

```

Possible site: 19

```

-1406-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2570(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA59773 GB:X85787 tasA [Streptococcus pneumoniae]
 Identities = 18/33 (54%), Positives = 28/33 (84%)

15 Query: 2 DVQSDENFAFKIFKVKAKGLSLDVFDKLVGRF 34
 + QSD+N F++FKV+K KG++LD FD+++GRF
 Sbjct: 320 EYQSDKNPFFFEVFKVSKTKGIALDPFDRIIGRF 352

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3909> which encodes the amino acid sequence <SEQ ID 3910>. Analysis of this protein sequence reveals the following:

Possible site: 56

20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2405(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 18/34 (52%), Positives = 25/34 (72%)

30 Query: 1 MDVQSDENFAFKIFKVKAKGLSLDVFDKLVGRF 34
 +DVQSDE+F FK+ KV K+KG+ L+ D+ V F
 Sbjct: 31 LDVQSDDEDFGFKVVKVLKSKGIVLNALDESVCGF 64

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

35 Example 1269

A DNA sequence (GBSx1346) was identified in *S.agalactiae* <SEQ ID 3911> which encodes the amino acid sequence <SEQ ID 3912>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

Possible site: 52

40 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.17 Transmembrane 169 - 185 (168 - 185)

----- Final Results -----

45 bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAC13546 GB:AF019629 putative fimbria-associated protein
 [Actinomyces naeslundii]
 Identities = 53/109 (48%), Positives = 75/109 (68%)

55 Query: 13 IPKINQDLPIYAGSEEDNLQRCVGHLEGISLPIGCSTHVLSCQRCMPAARLFADLDKM 72
 IP I+ DLP+Y G+ +D L +G+GHLEG SLP+GG T +V++G RG+ A +F +LDK+
 Sbjct: 93 IPSISLDLPVYHGTADDTLLKGLGHLEGTSLPVGGEGTRSVITGHRGLAEATMFTNLDKV 152

Query: 73 KKGDFYFVVTNLKETLAYQVDRIMVIEPSQLDAVSIEEDKDYVTLTCTP 121

-1407-

K GD V E L Y+V V+EP + +A+ +EE KD +TL+TCTP
 Sbjct: 153 KTGD SLIVEVFGEVLTYRVTS TKVVEPEETEALRVEEGDLLTLVTCTP 201

There is also homology to SEQ ID 3740 and to SEQ ID 3910.

- 5 SEQ ID 3912 (GBS194) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 177 (lane 2; MW 24kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1270

- 10 A DNA sequence (GBSx1347) was identified in *S.agalactiae* <SEQ ID 3913> which encodes the amino acid sequence <SEQ ID 3914>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 15 INTEGRAL Likelihood = -5.15 Transmembrane 880 - 896 (876 - 898)
 INTEGRAL Likelihood = -4.78 Transmembrane 24 - 40 (23 - 42)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3060(Affirmative) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8777> which encodes amino acid sequence <SEQ ID 8778> was also identified. Analysis of this protein sequence reveals the following:

25 Lipop: Possible site: -1 Crend: 8
 SRCFLG: 0
 McG: Length of UR: 20
 Peak Value of UR: 2.80
 Net Charge of CR: 5
 30 McG: Discrim Score: 10.81
 GvH: Signal Score (-7.5): -3.76
 Possible site: 29
 >>> Seems to have an uncleavable N-term signal seq
 Amino Acid Composition: calculated from 1
 ALOM program count: 2 value: -5.15 threshold: 0.0
 35 INTEGRAL Likelihood = -5.15 Transmembrane 867 - 883 (863 - 885)
 INTEGRAL Likelihood = -4.78 Transmembrane 11 - 27 (10 - 29)
 PERIPHERAL Likelihood = 7.58 531
 modified ALOM score: 1.53
 icml HYPID: 7 CFP: 0.306
 40 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3060(Affirmative) < succ>
 45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 LPXTG motif: 859-863

- 50 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8778 (GBS104) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 27 (lane 5; MW 95kDa).

GBS104-His was purified as shown in Figure 221, lane 9-10.

-1408-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1271

A DNA sequence (GBSx1348) was identified in *S.agalactiae* <SEQ ID 3915> which encodes the amino acid sequence <SEQ ID 3916>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -15.28    Transmembrane 257 - 273 ( 252 - 280)
10    INTEGRAL    Likelihood = -7.11    Transmembrane 19 - 35 ( 16 - 39)

----- Final Results -----
        bacterial membrane --- Certainty=0.7114 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC13546.GB:AF019629 putative fimbria-associated protein
    [Actinomyces naeslundii]
20    Identities = 79/178 (44%), Positives = 112/178 (62%), Gaps = 7/178 (3%)

Query: 65  RIALANAYNETLSRNPLL-----IDPFTSKQKEGLREYARMLEVHEQ--IGHVAIPSIGV 117
      ++  A+AYN+ LS  +L      +      K+  +YA +L+ + + + + IPSI +
Sbjct: 39  QVEQAHAYNDALSAGAVLEANNHVFTGAGSSKDSSLQYANILKANNEGLMARLKIPSISL 98

25    Query: 118 DIPIYAGTSETVLQKGSCHLEGTSLPVGGLSTHSLTAHRLPTARLFTDLNKVKKGQIF 177
      D+P+Y GT++  L KG GHLEGTSLPVGG  T SV+T HRGL  A +FT+L+KVK G
Sbjct: 99  DLPVYHGTADDTLLKGLGHLEGTSLPVGEGTRSVITGHRGLAEATMFTNLDKVKTGDSL 158

30    Query: 178 YVTNIKETLAYKVVSIIKVVDPTALSEVKIVNGKDYITLLTCTPYMINSHRLLVKGERI 235
      V   E L Y+V S KVV+P      +++  GKD +TL+TCTP  IN+HR+L+ GERI
Sbjct: 159 IVEVFGEVLTyrVTSTKVVPEETEALRVEEGKDLLTLVTCTPLGINTHRILLTGERI 216

```

There is also homology to SEQ ID 3740.

35 SEQ ID 3916 (GBS208) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 5; MW 35kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 8; MW 59.7kDa) and in Figure 160 (lane 5; MW 60kDa).

GBS208-GST was purified as shown in Figure 224, lane 7-8.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1272

A DNA sequence (GBSx1349) was identified in *S.agalactiae* <SEQ ID 3917> which encodes the amino acid sequence <SEQ ID 3918>. This protein is predicted to be a fimbria-associated protein. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -9.13    Transmembrane 265 - 281 ( 260 - 284)
50    ----- Final Results -----

```

-1409-

bacterial membrane --- Certainty=0.4652(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC13546 GB:AF019629 putative fimbria-associated protein
 [Actinomyces naeslundii]
 Identities = 96/265 (36%), Positives = 150/265 (56%), Gaps = 10/265 (3%)

10 Query: 41 QASHANINAFKEAVTKIDRVEINRRLELAYAYNASI-AGAKTNGEYPALKDPYSAEQKQA 99
 Q + + + A A R + ++E A+AYN ++ AGA P A +
 Sbjct: 15 QYNQSKVTADYSAQVDGARPDAKTQVEQAHAYNDALSAGAVLEANNHV---PTGAGSSKD 71

15 Query: 100 GVVEYARMLEVKEQ--IGHVVIIPRINQDIPIYAGSAEENLQRGVGHLEGTSLPVGGESTH 157
 ++YA +L+ + + + IP I+ D+P+Y G+A++ L +G+GHLEGTSLPVGGE T
 Sbjct: 72 SSLQYANILKANNEGLMARLKIPSISLDLPVYHGTADDTLLKGLHLEGTSLPVGEGTR 131

20 Query: 158 AVLTAHRGLPTAKLFTNLDKVTGDRFYIEHIGGKIAYQVDQIKVIAPDQLEDLYVIQGE 217
 +V+T HRGL A +FTNLDKV GD +E G + Y+V KV+ P++ E L V +G+
 Sbjct: 132 SVITGHRGLAEATMFTNLDKVKTGSLIVEVFGEVLTyrVTSTKVVPEETEALRVEEGK 191

25 Query: 218 DHVTLTCTPYMINSHRLLVRGKRI-PYVEKTVQKDSKTFRQQQYLTYAMWVVVGLILLS 276
 D ++TL+TCTP IN+HR+L+ G+RI P K + K + +A+ + GLI++
 Sbjct: 192 DLLTLVTCTPLGINTHRILLTGERIYPTPAKDLAAAGKRPDVPHPFWAVGLAAGLIVVG 251

Query: 277 LLIW---FKKTKQKKRRKNEKAASQ 298
 L +W + + K+R A+Q
 Sbjct: 252 LYLWRSGYAARAKERALARARAAQ 276

30 There is also homology to SEQ ID 3740.

SEQ ID 3918 (GBS209) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 4; MW 62kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 3; MW 37.2kDa).

GBS209-His was purified as shown in Figure 221, lane 8.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1273

A DNA sequence (GBSx1350) was identified in *S.agalactiae* <SEQ ID 3919> which encodes the amino acid sequence <SEQ ID 3920>. Analysis of this protein sequence reveals the following:

40 Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -9.66 Transmembrane 281 - 297 (276 - 300)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.4864(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:BAB04080 GB:AP001508 unknown [Bacillus halodurans]
 Identities = 45/141 (31%), Positives = 63/141 (43%), Gaps = 20/141 (14%)

55 Query: 153 TGELDLLKVGVDGDTKKPLAGVVFELYEKNGRTPIRVKNGVHSQDIDAAKHLETDSSGHI 212
 TG L++ KV D DT + L G F LY+ G IR LET G
 Sbjct: 1084 TGSLEVTKV--DADTGEVLQCATFTLYDSEGEFAIRT-----LETGEDGKA 1127

-1410-

Query: 213 RISGLIHGDYVLKEIETQSGYQIGQAETAVTIEKSKTVTVTTIENKKVPTPKVPSRGGL-I 271
 L++GDY+LKE GY +G +T + VT+EN+K +V + G + +
 Sbjct: 1128 TFWNLLYGDYLLKEDSAPEGYLVGINDTQRTIDTVLHEVTVENEKSDINRVSAVGAVQL 1187

Query: 272 PKTGEQQAMALVIIGGILIAL 292
 K E+ +L G L AL
 Sbjct: 1188 QKVDEETGESL---QGALFAL 1205
 Identities = 64/259 (24%), Positives = 113/259 (42%), Gaps = 48/259 (18%)

Query: 16 GTMFGISQT---VLAQETHQLTIVHLEARDIDRPNP---QLBIAPKE-GTPIEGVLYQL 67
 G + GI+ T + H++T+ + E DI+R + QL+ +E G ++G L+ L
 Sbjct: 1147 GYLVGINDTQRTIDTVLHEVTVEN-EKSDINRVSAVGAVQLQKVDEETGESLQGALFAL 1205

Query: 68 YQLKSTEDGDL LAHWNSLTITELKKQAQOVFEATTNQQKATFNQLPDGIYYGL---AV 123
 Q E +TI E++ + + A + + G F +L + Y L V
 Sbjct: 1206 QQKVDDE-----FVTIAEMETDEEGIVFAGSLEPGDYQFVELNAPVGYKLDETPVV 1256

Query: 124 KAGEKNRNVSAFLVDLSEDKVIYPKIIWSTGELDLLKVGVDGDTKKPLAGVVFELYEKNG 183
 E++R + ++L ++ + P G + L+KV D D L G F L + G
 Sbjct: 1257 FTVEEDRTET---IELQKENHLIP-----GSVQLVKVDAD-DAANTLEGAEFTLLDGE 1306

Query: 184 RTPIRVKNGVHSDIDAAKHLETSSSGHIRISGLIHGDYVLKEIETQSGYQIGQAETAVT 243
 V+ G L TD +G + ++ L G+Y E + +GY++ T
 Sbjct: 1307 NV---VQEG-----LTTDENGQVVVDLKPGEYQFVETKAPAGYELEATPIGFT 1352

Query: 244 IEKS--KTVTVTTIENKKVP 260
 IE++ + TV +EN +P
 Sbjct: 1353 IERNQQEVATVAVENHLIP 1371

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3920 (GBS52) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 4; MW 30.5kDa).

GBS52-His was purified as shown in Figure 192, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1274

A DNA sequence (GBSx1351) was identified in *S.galactiae* <SEQ ID 3921> which encodes the amino acid sequence <SEQ ID 3922>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -6.26 Transmembrane 554 - 570 (551 - 575)
 INTEGRAL Likelihood = -0.16 Transmembrane 34 - 50 (34 - 50)

----- Final Results -----
 bacterial membrane --- Certainty=0.3506(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8779> which encodes amino acid sequence <SEQ ID 8780> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0
 McG: Discrim Score: -5.81
 GvH: Signal Score (-7.5): -1.92
 Possible site: 37
 >>> Seems to have a cleavable N-terminal signal sequence
 ALOM program count: 2 value: -6.26 threshold: 0.0

-1411-

INTEGRAL Likelihood = -6.26 Transmembrane 527 - 543 (524 - 548)
 PERIPHERAL Likelihood = 5.36 194
 modified ALOM score: 1.75

5 *** Reasoning Step: 3

----- Final Results -----

10 bacterial membrane --- Certainty=0.3506(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 521-525

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAA57459 GB:X81869 orf2 [Lactobacillus leichmannii]
 Identities = 140/505 (27%), Positives = 220/505 (42%), Gaps = 94/505 (18%)

Query: 102 GEVISNYAKLGDNVKGLQGVQFKRYKVKTDI-----SVDELKKLTTVEAADAKVGTILEE 156
 GE+++++ G L GV FK Y V S D + T +DAK L

20 Sbjct: 58 GEIMNDFGGTG-----LNGVTFKAYNVTDHYLSLRKSGDSAQDAVTAIQSDAKSDNLPS 112

Query: 157 --GVSLPQKTNAQGLVVDAL-----DSKSNVR-PLYVEDLKNSPSNITKAYAVPFV 204
 G ++ +T A D + DS N + YL+VE +SP+++T+ A P V

25 Sbjct: 113 YAGSAIATETTATSKGEDGIAAFDNLNLKDSGNYQTYLFVET--DSPTDVTQQ-AAPIV 169

Query: 205 LELPVANSTGTGFLS-EINIYPKNVVTDEPKTKDKVKKLGQDDAGYTI-----G 252
 L +P+ ++ T ++ +I IYPKNV + P T KD+ + + D T+ G

30 Sbjct: 170 LTMPIYKTSOTSAINHDIQIYPKNVKST-PIT-KDLDEASKDLAVTLPDGSTIYNAQYG 227

Query: 253 EEFKWLKSTIPANLGDYEFKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVD 312
 + F + + +P N+ D + F + DK G+ + L + YT+++

35 Sbjct: 228 KSGFYNITVNVPNWIKDKDTPNVVDKPDGTI---DIDASTVSDGLTKSTDYTVNK---- 280

Query: 313 NQNTLKITFKPEKFKEIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVL 372
 N ++ FK + L G +L I +T+ A

40 Sbjct: 281 KONGYQVVFKTTS--AAVQALACKSLT-----ITYKATLTNNATP 318

Query: 373 GKAIENFELQYDHTPDKADNPKNPSPRKPPEVHTGGKRFVKKDSTETQTLGGAEFDLA 432
 KAI NT L + + S P P ++TGG +FVKDS +TL GAEF L+

45 Sbjct: 319 DKAIGNATLSIGNGTNIT-----STPANGPRIYTGAQFVKKDSQSNKTLAGAELQVLK 373

Query: 433 --SDGTAVKWTDALIKANTKNYIAGEAVTGQPIKLKSHDTGTFEIKGLAYAVDANAEGT 490
 S+G V + + N A EA T S +G +KGL+Y ++ +

50 Sbjct: 374 VDSNGNIVSYATQASDGSYTNDSATEATT-----YTSANGLVALKGLSY---SDKLDS 425

Query: 491 AVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPN 550
 +Y L E +AP+GY D ++F+++Q S+ D+ TI N K +P+

55 Sbjct: 426 GESYALLEIQAPDGYAKLDSVPKFSITQGSF-----GDSNKITIDNTKEGLLPS 474

Query: 551 TGGIGTAIFVAIGAAMFAVKGGMK 575
 TGG G IF+AIG +M A G K

60 Sbjct: 475 TGGKGIYIFLAIGIVIMIVAFGGYK 499

No corresponding DNA sequence was identified in *S.pyogenes*.

55 SEQ ID 8780 (GBS80) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 6; MW 56.8kDa).

The GBS80-His fusion product was purified (Figure 104A; see also Figure 194, lane 5) and used to immunise mice (lane 1+2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 104B), FACS (Figure 104C), and in the *in vivo* passive protection assay (Table III). These tests confirm
 60 that the protein is immunoaccessible on GBS and that it is an effective protective immunogen.

-1412-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1275

A DNA sequence (GBSx1352) was identified in *S.agalactiae* <SEQ ID 3923> which encodes the amino acid sequence <SEQ ID 3924>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4043(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1276

A DNA sequence (GBSx1353) was identified in *S.agalactiae* <SEQ ID 3925> which encodes the amino acid sequence <SEQ ID 3926>. This protein is predicted to be MsmR. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.01 Transmembrane 75 - 91 (75 - 92)
----- Final Results -----
bacterial membrane --- Certainty=0.1404(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9679> which encodes amino acid sequence <SEQ ID 9680> was also identified.

SEQ ID 3926 (GBS360) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 9; MW 74kDa).

GBS360-GST was purified as shown in Figure 216, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1277

A DNA sequence (GBSx1354) was identified in *S.agalactiae* <SEQ ID 3927> which encodes the amino acid sequence <SEQ ID 3928>. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.1762(Affirmative) < succ>

-1413-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3929> which encodes the amino acid sequence <SEQ ID 3930>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1640 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 93/98 (94%), Positives = 96/98 (97%)
 Query: 1 MDKIIKSISASGAFRSYVLDTSTETVKLAQEKHHTLSSSTVALGRTLIANQILAANQKGDS 60
 MDKIIKSI+ SGAFR+YVLDTSTETV LAQEKH+TLSSSTVALGRTLIANQILAANQKGDS
 20 Sbjct: 1 MDKIIKSIAQSGAFRAYVLDTSTETVALAQEKHNTLSSSTVALGRTLIANQILAANQKGDS 60
 Query: 61 KITVKVIGDSSFGHIISVADTKGHVKGYIQNTGVDIKK 98
 KITVKVIGDSSFGHIISVADTKGHVKGYIQNTGVDIKK
 Sbjct: 61 KITVKVIGDSSFGHIISVADTKGHVKGYIQNTGVDIKK 98

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1278

A DNA sequence (GBSx1355) was identified in *S.agalactiae* <SEQ ID 3931> which encodes the amino acid sequence <SEQ ID 3932>. Analysis of this protein sequence reveals the following:

30 Possible site: 17
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 35 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:AAC98436 GB:L29324 unknown [Streptococcus pneumoniae]
 Identities = 34/48 (70%), Positives = 39/48 (80%)
 Query: 1 MQEVLIIARENHQVTHEHVSILLTCVQELIVEVNQTQPLSREFREKYM 48
 + EV IIA+ NHQVTHEHVSILLTC+QELI EV +T PLS +F KYM
 45 Sbjct: 70 VHEVFIIAKTNHQVTHEHVSILLTCIQELIKEVEKTGPLSEDFCNKYM 117

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1279

50 A DNA sequence (GBSx1356) was identified in *S.agalactiae* <SEQ ID 3933> which encodes the amino acid sequence <SEQ ID 3934>. This protein is predicted to be TnpA (orfB). Analysis of this protein sequence reveals the following:

-1414-

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.5248 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9907> which encodes amino acid sequence <SEQ ID 9908>
 10 was also identified. A further related GBS nucleic acid sequence <SEQ ID 9677> which encodes amino acid
 sequence <SEQ ID 9678> was also identified. A further related GBS nucleic acid sequence <SEQ ID
 10911> which encodes amino acid sequence <SEQ ID 10912> was also identified.

There is homology to SEQ ID 1336.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 15 vaccines or diagnostics.

Example 1280

A DNA sequence (GBSx1357) was identified in *S.agalactiae* <SEQ ID 3935> which encodes the amino
 acid sequence <SEQ ID 3936>. Analysis of this protein sequence reveals the following:

20 Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.4489 (Affirmative) < succ>
 25 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB64982 GB:U43834 Ydr540cp [Saccharomyces cerevisiae]
 Identities = 93/171 (54%), Positives = 121/171 (70%), Gaps = 3/171 (1%)
 30 Query: 1 MRVYENKEELKKEISKTFEKYIMEFNNIPENLKDKRIDEVDRTTPAANLSYQVGWTNLVLK 60
 MR Y +K+ELK+EI K +EKY EF I E+ KD++++ VDRTP+ NLSYQ+GW NL+L+
 Sbjct: 1 MREYTSKKELKEEIEKKYKYDAEFETISESQKDEKVETVDRTPSENLSYQLGWNLLLE 60
 35 Query: 61 WEEDERKGLQVKTPSDKFKWNQLGELYQWFTDTYAHLSLQELKAKLNENINSIYAMIDLL 120
 WE E G V+TP+ +KWN LG LYQ F Y S++E +AKL E +N +Y I L
 Sbjct: 61 WEAKEIAGYNVETPAPGYKWNNLGGLYQSIFYKYGYSIKEQRAKLREAVNEVYKWISTL 120
 40 Query: 121 SEEEELFEAHMRKWADEATKTATWEVYKFIHVNTVAPFGTFRTKIRKWKKIV 171
 S++ELF+A RKW AT A W VYK+IH+NTVAPF FR KIRKWK++V
 Sbjct: 121 SDDELFPQAGNRKW--ATTKAMWPVYKWIHINTVAPFTNFRGKIRKWKRLV 168

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 45 vaccines or diagnostics.

Example 1281

A DNA sequence (GBSx1358) was identified in *S.agalactiae* <SEQ ID 3937> which encodes the amino
 acid sequence <SEQ ID 3938>. Analysis of this protein sequence reveals the following:

50 Possible site: 28
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -3.45 Transmembrane 10 - 26 (2 - 26)

-1415-

----- Final Results -----

5 bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8781> which encodes amino acid sequence <SEQ ID 8782> was also identified. Analysis of this protein sequence reveals the following:

10 Lipop: Possible site: -1 Crend: 6
 McG: Discrim Score: 8.80
 GvH: Signal Score (-7.5): -3.94
 Possible site: 28
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 1 value: -3.45 threshold: 0.0
 15 INTEGRAL Likelihood = -3.45 Transmembrane 7 - 23 (2 - 26)
 PERIPHERAL Likelihood = 10.40 69
 modified ALOM score: 1.19
 *** Reasoning Step: 3
 20 ----- Final Results -----
 bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000(Not Clear)

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA68889 GB:Y07615 acid phosphatase [Haemophilus influenzae]
 Identities = 112/245 (45%), Positives = 148/245 (59%), Gaps = 10/245 (4%)
 30 Query: 5 MKKVLVSSLLVLGITITLQTVVEAKGPKVAYTQEGMTALSDTNKDKVTTISIDEIQKSLE 64
 MK V+ S++ L +T V G YTQ G A + + IS+D+I++SLE
 Sbjct: 1 MKNVMKLSVIAL---LTAAAVPAMAGKTEPYTQSGTNAREMLQEQAIIHWISVDQIKQSLE 57
 35 Query: 65 GKKPITVSFDIDDTLLFSSQYFQYQKEYVTPGSDFLHKQKFWDLVAKRGDQDSIPKEYA 124
 GK PI VSEFDIDDT+LFSS F +G++ +PG D+L Q FW+ V D+ SIPK+ A
 Sbjct: 58 GKAPINVSEFDIDDTVLFSPPCFYHGGQKFSPGKHLYLNQDFWNEVNAGCDKYSIPKQIA 117
 Query: 125 KKLIAMHQKRGDKIVFITGRTRGSMYKEGEVDKTAALAKDFKLDKPIAVNYTGDPKPKP 184
 LI MHQ RGD++ F TGR T G+VD L K F + V + G + ++
 40 Sbjct: 118 IDLINMHQARGDQVYFFTGR T-----AGKVDGVTPILEKTFNIIKNMHFVEFMGSR-ERT 170
 Query: 185 YKYDKSYIYIKKYGSDIHYGSDDDDIHAAREAGARPIRILRAPNSTNLPLPEAGGYGEEVL 244
 KY+K+ I + IHYGSDDD+ AA+EAG R IR++RA NST P+P GGYGEEVL
 45 Sbjct: 171 TKYNKTPAILSHKVSIIHYGSDDDVLAAKEAGVRGIRLMRAANSTYQPMPTLGGYGEEVL 230
 Query: 245 ENSAY 249
 NS+Y
 Sbjct: 231 INSSY 235

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3939> which encodes the amino acid sequence <SEQ ID 3940>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have an uncleavable N-term signal seq
 55 INTEGRAL Likelihood = -3.98 Transmembrane 6 - 22 (4 - 25)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 60 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA68889 GB:Y07615 acid phosphatase [Haemophilus influenzae]

-1416-

Identities = 105/237 (44%), Positives = 141/237 (59%), Gaps = 10/237 (4%)

5 Query: 9 LFTVSFCGIIALPVEASGPKVPYTOEGITA--ISNQATVKLISIADIASSLEGQKPITVS 66
 L ++ A+P A G PYTQ G A + + + IS+ I SLEG+ PI VS
 Sbjct: 7 LSVIALLTAAAVPAMA-GKTEPTQSGTNAREMLQEQAIIHWISVDQIKQSLEGKAPINVS 65

10 Query: 67 FDIDDTLLFTSQYFQYQYKEYITPGSFDFLHKQKFWDLVAKRGDQDSIPKEYAKQLIAMHQ 126
 FDIDDT+LF+S F +G++ +PG D+L Q FW+ V D+ SIPK+ A LI MHQ
 Sbjct: 66 FDIDDTVLFFSSPCFYHQKQKFSQKHDYLNQDFWNEVNAGCDKYSIPKQIAIDLINMHQ 125

15 Query: 127 KRQDKIVFITGRTRGSMYKKGEIDKTAKSLAKDFKLDKPIAINYTGDKAVKPYQYDKTTY 186
 RGD++ F TGRF G++D L K F + + + G + + +Y+KT
 Sbjct: 126 ARGDQVYFFTGRF-----AGKVDGVTFPILEKTFNKNMHPVEFMGSRE-RTTKYNKTPA 178

Query: 187 IKKNGSQIHYGSDSDEDINAAREAGARPILRAPNSTNLPLPKAGGYGEEVLNSAY 243
 I + IHYGSDS+D+ AAKEAG R IR++RA NST P+P GGYGEEVL NS+Y
 Sbjct: 179 IISHKVSIIHYGSDDDVLAAKEAGVRGIRLMRAANSTYQPMPTLGGYGEEVLINSSY 235

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 196/245 (80%), Positives = 216/245 (88%), Gaps = 2/245 (0%)

Query: 5 MKKVLVSSLLVLGITITLQTVVEAKGPKVAYTOEGMTALSDTNKDKVTTISIDEIQKSLE 64
 MKK S L + + VEA GPKV YTOEG+TA+S N+ V ISI +I SLE
 25 Sbjct: 1 MKKEFTSILFTVSFCGIIALPVEASGPKVPYTOEGITAIS--NQATVKLISIADIASSLE 58

Query: 65 GKKEPITVSFDIDDTLLFSSQYFQYQYKEYVTPGSFDFLHKQKFWDLVAKRGDQDSIPKEYA 124
 G+KEPITVSFDIDDTLLF+SQYFQYQYKEY+TPGSFDFLHKQKFWDLVAKRGDQDSIPKEYA
 30 Sbjct: 59 GQKEPITVSFDIDDTLLFTSQYFQYQYKEYITPGSFDFLHKQKFWDLVAKRGDQDSIPKEYA 118

Query: 125 KKLIAHQKRGDKIVFITGRTRGSMYKEGEVDKTAALAKDFKLDKPIAVNYTGDKPKKP 184
 K+LIAHQKRGDKIVFITGRTRGSMYK+GE+DKTAK+LAKDFKLDKPIA+NYTGDK KP
 35 Sbjct: 119 KQLIAHQKRGDKIVFITGRTRGSMYKKGEIDKTAKSLAKDFKLDKPIAINYTGDKAVKP 178

Query: 185 YKYDKSYIYIKKYSIDIHYGSDDDIHAAREAGARPILRAPNSTNLPLPEAGGYGEEVL 244
 Y+YDK+YIYIKK GS IHYGSDS+DI+AA+EAAREAGARPILRAPNSTNLPLP+AGGYGEEVL
 40 Sbjct: 179 YQYDKTYIYIKKNGSQIHYGSDSDEDINAAREAGARPILRAPNSTNLPLPKAGGYGEEVL 238

Query: 245 ENSAY 249
 ENSAY
 Sbjct: 239 ENSAY 243

SEQ ID 8782 (GBS100) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 5; MW 28kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 2; MW 53kDa).

45 The GBS100-GST fusion product was purified (Figure 106A; see also Figure 197, lane 4) and used to immunise mice (lane 1 product; 9.9µg/mouse). The resulting antiserum was used for Western blot (Figure 106B), FACS (Figure 106C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1282

A DNA sequence (GBSx1359) was identified in *S.agalactiae* <SEQ ID 3941> which encodes the amino acid sequence <SEQ ID 3942>. Analysis of this protein sequence reveals the following:

55 Possible site: 19
 >>> Seems to have no N-terminal signal sequence

-1417-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3288(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 1283

A DNA sequence (GBSx1360) was identified in *S.galactiae* <SEQ ID 3943> which encodes the amino acid sequence <SEQ ID 3944>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4004(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9675> which encodes amino acid sequence <SEQ ID 9676> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25

>GP:BAB04406 GB:AP001509 RNA methyltransferase [Bacillus halodurans]
 Identities = 198/452 (43%), Positives = 300/452 (65%)

30

Query: 12 KRKIMLHKNDIETEISDISHEGMGIKVDGFFVENALPGEIIMRVLKLKRIGYGK 71
 K++ ++KND++E I D++H+G G+AKVDG+ F+ ALPGE +K +V+K++K G+G+
 Sbjct: 3 KQAPVKNKNDVVEVTIEDLTHDGAGVAKVDGYALFIPKALPGERLKAKVVKVKKGYGFR 62

35

Query: 72 VEEYLTTSFHRNEGLDYTYLRTGIADLGHLYEQQLLFKQKQVADNLYKIAHISDVLVEP 131
 V + SP R E ++ G L H++Y+ QL +KQKV D L +I I+ V V P
 Sbjct: 63 VLNMTIASPDRVEAPCPVFNQCGGQQLQHMSYDAQLRYKQKQVQDVLERIGKITAVTVRP 122

40

Query: 132 TLGMTIPLAYRNKAQVPVRRVDGQLETGFFRKNSHTLVSIEDYLIQEKEIDALINFTRD 191
 T+GM P YRNKAQVPV +G L GF+++ SH ++ +++ +IQ +E D +I ++L
 Sbjct: 123 TIGMNEPWRYRNKAQVPVGEREGGLIAGFYQERSHRIIDMDECMIQHEENDKVIRQVKEL 182

45

Query: 192 LRKFDVKPYDEEQQSGLIRNLVVRGHYTGQMLVLVTTTRPKIFRIDQMIEKLVSAPFSV 251
 R+ ++ YDEE+ G +R++V R G TG++M+VL+T ++ +IE++ A P V
 Sbjct: 183 ARELGIRGYDEEKHRTLRHVVARVYKNTGEIMVVLITRGEELPHKKTILIERIHKAIPHV 242

50

Query: 252 VSIMQNINDRNSNVIFGKEFRTLYGSDTIEDQMLGNTYAIASQSFYQVNTMAEKLYQKA 311
 SI+QN+N + +NVIFG + + L+G + I D + +AISA+SFYQVN E + LY +A
 Sbjct: 243 KSIVQNVNPKRTNVIFGDKTKVLWGEEIYDTIGDIKFAISARSFYQVNPEQTKVLYDQA 302

55

Query: 312 IDFSDLNSEDIVIDAYSGIGTIGLSVAKQVKHVYGVVEVVEKAVSDAKENATRNGITNSTY 371
 ++F++L + VIDAY GIGTI L +A+Q KHVYGVVE+V +A+SDAK NA NG N +
 Sbjct: 303 LEFANLTGSETVIDAYCGIGTISLFLAQQAQKHVYGVVEIVPEAISDAKRNARLNGFANVQF 362

Query: 372 VADSAENAMAKWLKEGIKPTVIMVDPKRLTESFVYSAAQTKADKITYISCNSATMARD 431
 AE M W +G++ VI+VDPKRG E+ + + K D++ Y+SCN AT+ARD
 Sbjct: 363 AVGDAEKVMPWWYAQGVVRADVIVVDPKRGCEALLKTLNMPKDRVVYVSCNPATLARD 422

Query: 432 IKLFEELGYHLVKIQPVDLFPMTTHHVECVALL 463
 +++ E+ GY +QPVD+FP T H+E VA+L
 Sbjct: 423 LRVLEDGGYETKDVQPVDMFPWTTTHIESVAVL 454

-1418-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3945> which encodes the amino acid sequence <SEQ ID 3946>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1262(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 332/454 (73%), Positives = 387/454 (85%)

Query: 12 KRKIMLHKNDIETEISDISHEGMGIKVDGFFVFFVENALPGEIIMRVLKLRKRIGYGK 71
KR ML KNDII+ ISD+SHEG G+AK DGFVFFV+NALP E+I MRVLK+ K G+GK
Sbjct: 8 KRIRMLKNDIIQVAISDLSEHGACVAKHDGFFVFFVDNALPEEVIDMRVLKVNKNSGFGK 67

Query: 72 VEEYLTTSPHRNEGLDYTYLRTGIADLGHLYEQQLLFKQKQVADNLYKIAHISDVLVEP 131
VE Y S RN ++ TYLRTGIADLGHLYE QL FK+KQV D+LYKIA ISDV VE
Sbjct: 68 VEAYHYLSSARNADVNLTYLRTGIADLGHLYEQQLLFKQKQVADNLYKIAHISDVLVEP 127

Query: 132 TLGMTIPLAYRNKAQVPVRRVDGQLETGFFRKNSHTLVSIEDYLIQEKEIDALINFTRDL 191
T+GMT PLAYRNKAQVPVRRV+GQLETGFFRK+SH L+ I DY IQ+KEID LINFTRDL
Sbjct: 128 TIGMTEPLAYRNKAQVPVRRVNGQLETGFFRKNSHTLVSIEDYLIQEKEIDALINFTRDL 187

Query: 192 LRKFDVVKPYDEEQSGILRNLVVRRGHYTGQMLVLVLTTRPKIFRIDQMIEKLVSAPFSV 251
LR+FD+KPYDE +Q+GL+RN+VVRGHY+G++MLVLVLTTRPK+FR+DQ+IEK+V AFP+V
Sbjct: 188 LRRFDIKPYDETEQTGLLRNIVVRRGHYSGEMMLVLVLTTRPKVFRVDQVIEKIVEAPFAV 247

Query: 252 VSIMQNINDRNSNVIFGKEPRTLYGSDTIEDQMLGNTYAIQAQSFYQVNTMAEKLYQKA 311
VSI+QNIND+N+N IFGK+F+TLYG DTI D MLGN YAIQAQSFYQVNT MAEKLYQ A
Sbjct: 248 VSIIQNINDKNTNAIFGKDFKTLYGKDTITDSMLGNNTYAIQAQSFYQVNTVMAEKLYQTA 307

Query: 312 IDFSDLNSEDIVIDAYSGIGTIGLSVAKQVKHVGVEVVEKAVSDAKENATRNGITNSTY 371
I FSDL+ +DIVIDAYSGIGTIGLS AK VK VYGVEV+E AV DA++NA NGITN+ +
Sbjct: 308 IAFSLSKDDIVIDAYSGIGTIGLSFAKTVKAVYGVEVIEAAVRDAQQAALNGITNAYF 367

Query: 372 VADSAENAMAKWLKEGKPTVIMVDPKRLTESFVYSAAQTKADKITYISCNSATMARD 431
VAD+AE+AMA W K+GIKP+VI+VDPKRLTESF+ ++ KITY+SCN ATMARD
Sbjct: 368 VADTAEHAMATWAKDGKPSVILVDPKRLTESFIQASVAMGPQKITYVSCNPATMARD 427

Query: 432 IKLFEELGYHLVKIQPVDLFPMTTHVECVALLVK 465
IK ++ELGY L K+QPVDLFP THHVECV LL+K
Sbjct: 428 IKRYQELGYKLAKVQPVDFPQTHHVECVVLLIK 461

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1284

A DNA sequence (GBSx1361) was identified in *S.galactiae* <SEQ ID 3947> which encodes the amino acid sequence <SEQ ID 3948>. This protein is predicted to be PSR protein. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -12.15 Transmembrane 135 - 151 (127 - 155)

----- Final Results -----

bacterial membrane --- Certainty=0.5861(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1419-

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB76822 GB:AJ276232 PSR protein [Enterococcus faecalis]
Identities = 143/409 (34%), Positives = 206/409 (49%), Gaps = 56/409 (13%)

5   Query: 48 QRRTESPP--TNSYYEFPYSDSYQDDDFYSEPOLTSQGLPIYQEERAPKKKKQARKEK 105
      + R E P      S E Y DSY +D      T G      ++ P+ KK + K+K
      Sbjct: 31 EHREEEPEELAESLQEPVYEDSYTEDSRRSERRHQTDSGGG-NGSDQPPRGKKDKKPKKK 89

10  Query: 106 QRVKVMAPPPPKAITPPRKKKKFKGFLKFIGIILLIVLSGMVFMFVKMRD VNNGKSHYS 165
      RKK K K F K++ I+L+++ +      MF+KG      + S
      Sbjct: 90 -----RKKSKTKRFFKWL VILLILLFAYSTVMFLKGKSAAEHDDS-LP 131

      Query: 166 PAIIEDFKGDVAVDGT-NILILGSDKRVSRSTDA RTDTIMVANVGKNKDKVMVSFMRD 224
      +E F G + +G NILILGSD R +      R DTIMV +      K K++SFMRD
15  Sbjct: 132 QEKVETFNQVSSNGAKNIIILGSDTRGEDAG---RADTIMVLQLNGPSKKPKLISFMRD 188

      Query: 225 LLVNIPNYSTEGYYDMKLNASFNLGEQDNHKGAEYVRQTLKNHFDIDIKYYVMVDFETFA 284
      V+IP      G      K+NA++ G      GAE VR+TLK +F++D KY Y VDF++F
20  Sbjct: 189 TFVDIP-----GVGPNKINAAYAG-----GAE LVRET LKQNFNLDTKYYAKVDFQSF 237

      Query: 285 DAIDTLFPNGVKINAKFGLVGGQSADSVKVPDDL RMKNGVVP SQIKVGIQYMDGRTLLN 344
      +D++FP GVKI+A+ L      + D V      I+ G Q MDG LL
      Sbjct: 238 KIVDSMFPGVKIDAESL----NL DGV-----IEKGQQVMDGHVLLQ 277

25  Query: 345 YARFRKDDGDFGRTQRQQQVMRAIVSQIKDPRRLFTGSAAGKAYALTSNLSYSFVL 404
      YARFR D++GDFGR +RQQQVM A++SQ+K+P L      ++GK      S+++ SF+LT
      Sbjct: 278 YARFRMDEBGDFGRVRRQQQVMSAVMSQMNPMILLRTPESLGKLVGYMSTDVPVSFMLT 337

30  Query: 405 DGIPILSDAKNGIKQMTIPREGDWDDYDQYGGQGLTIDFAKYKKILKK 453
      +G +L K G++ +++P W      Y G L +D K      ++K
      Sbjct: 338 NGPSLLIKGKTGVESLSVPVDSWNFGESSYAGSILEVDEQKNADAIEK 386

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3949> which encodes the amino acid sequence <SEQ ID 3950>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 49

      >>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -7.96 Transmembrane 159 - 175 ( 152 - 180)

40  ----- Final Results -----
      bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

45 The protein has homology with the following sequences in the databases:

```

>GP:CAB76822 GB:AJ276232 PSR protein [Enterococcus faecalis]
Identities = 140/345 (40%), Positives = 195/345 (55%), Gaps = 41/345 (11%)

50  Query: 140 PRSQK----RKHKKKGCMKWFFNIGLLMTVLMGLMFAGKGVFDISTNKANYKPAVSQ 195
      PR +K      +K +KK K FF L +LL+ +      +MF KG      + + + V +
      Sbjct: 78 PRGKKDKKPKKKRKKSKTKRFFKWL VILLILLFAYSTVMFLKGKSAAEHDDSLPQEKV-E 136

      Query: 196 AFDGQETQDGT-NILILGSDQRVTQGSTDA RTDTIMVVNVGNHAKKIKMVSFMRD TLINI 254
      F+G ++ +G NILILGSD T+G R DTIMV+ +      +KK K++SFMRDT ++I
55  Sbjct: 137 TFNGVKSSNGAKNIIILGSD---TRGEDAGRADTIMVLQLNGPSKKPKLISFMRDTFVDI 193

      Query: 255 PGYSYNDNSYDLKLNASFNLGEQEDHHGAEYVRRALKHNFIDIDIKYYVMVDFETFAE AID 314
      PG N      K+N+A+ G      GAE VR LK NF++D KY Y VDF++F + +D
60  Sbjct: 194 PGVGPN-----KINAAYAG-----GAE LVRET LKQNFNLDTKYYAKVDFQSF EKIVD 241

      Query: 315 TLFPNGVKIDAKFATVGGVAVDSEVPDDL RMKNGVVPNQTTIEVGEQRM DGR TLLNYARF 374
      ++FP GVKIDA+ +      + +D V+      IE G+Q MDG LL YARF
      Sbjct: 242 SMFPKGVKIDAESK----LNLDGVD-----IEKGQQVMDGHVLLQYARF 281

```

-1420-

Query: 375 RKDDEGDFGRTVRQQQVMSAVMSQIKDPTKLFTGSAAGKIYALITSTNVSFPFVVKNGVS 434
 R D+EGDFGR RQQQVMSAVMSQ+K+P L ++GK+ ST+V F++ NG S
 Sbjct: 282 RMDEEGDFGRVRRQQQVMSAVMSQIMNPMTLRLTPESLGKLVGYMSTDPVPSFMLTNGPS 341

5 Query: 435 VLGSGKNGVEHVTIPENGDWVDEYDMYGGQALYIDFDKYQKTLAK 479
 +L GK GVE +++P W Y G L +D K + K
 Sbjct: 342 LLIKPKTGVESLSVPVDSWNFGESSYAGSILEVDEQKNADAIEK 386

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 273/486 (56%), Positives = 340/486 (69%), Gaps = 32/486 (6%)

Query: 1 MSRNNGYQLNHHEELRYNYLLKNIHYLNEREKMEFQYLHYKKTAVRPQRRTESPTNSYY 60
 M++ G L+HHEELRY YLL+N+ YL+E EK EF +L K R ++ S
 Sbjct: 1 MTKYPMGGLSHHEELRYFYLLRNLSYLSENEKKEFAFLKSKLEIGRAYAPSKQHYRKSQR 60

15 Query: 61 EEPY-SDSY-----QDDDFYSEPOLTSQGLPIYQEEERAPKKKQRRARKEKQVRKV 110
 +EPY D YY +DDD + GLPIY +E KK K R +
 Sbjct: 61 QEPYFEDDYNDYSPNDLLEDDDVNHDSSFPYGLPIYPKEDRYLNKKT---KLTARRPI 117

20 Query: 111 MAPFP-----PKAITPPRKKKK-FKGFLKFIGIILLIIVLSGMVFMFVK 152
 AP P P++ KKK K F +G++L+ VL G+ MF K
 Sbjct: 118 DAPQPIDEDDAFLTESVARCALPRSQRKHKKKGCMKWFFNILGLLLMTVLMGLGLMFAK 177

25 Query: 153 GMRDVNNGKSHYSPAIEDFKGKDAVDGNTNLIILGSDKRVSERSTDARTDTIMVANVGNK 212
 G+ D++ K++Y PA+ + F G++ DGTNLIILGSD+RV++ STDARTDTIMV NVGN
 Sbjct: 178 GVFDISTNKANYKPAVSQAFDQGTQDGTNLIILGSDQRVVTQGSTDARTDTIMVVNVGNH 237

30 Query: 213 DNKVMVMSFMRDLLVNIIPNYS-TEGYDMKLNASFNLGEQDNHKGAEYVRQTLKNHFDID 271
 K+KMVSFMRD L+NIP YS + YD+KLN++FNLGEQ++H GAERYVR+ LK++FDID
 Sbjct: 238 AKKIKMVMSFMRDTLINIPGYSYNDNSYDLKLNASFNLGEQEDHHGAERYVRALKHNFID 297

35 Query: 272 IKYYVMVDFETFAAIDTLFPNGVKINAKFGLVGGQSADSVKVPDDLRLMKNGVVPVSKIK 331
 IKYYVMVDFETFAAIDTLFPNGVKI+AKF VGG + DSV+VPDDLRLMKNGVVP+Q I+
 Sbjct: 298 IKYYVMVDFETFAAIDTLFPNGVKIDAKFATVGGVAVDSEVPDDLRLMKNGVVPNQITIE 357

40 Query: 332 VGIQYMDGRTLNLNARFRKDDGDFGRTRQQQVMRAIVSQIKDPRRLFTGSAAGKAYA 391
 VG Q MDGRTLNLNARFRKDD+GDFGRTRQQQVM A++SQIKDP +LFTGSAAGK YA
 Sbjct: 358 VGEQRMGRTLNLNARFRKDDGDFGRTRQQQVMSAVMSQIKDPTKLFTGSAAGKIYA 417

45 Query: 392 LTSSNLSYSFVLTDGIPILSDAKNGIKQMTIPREGDWDDYDQYGGQGLTIDFAKYKKIL 451
 LTS+N+S+ FV+ +G+ +L KNG++ +TIP GDWVD+YD YGGQ L IDF KY+K L
 Sbjct: 418 LTSTNVSFPFVVKNGVSVLGSGKNGVEHVTIPENGDWVDEYDMYGGQALYIDFDKYQKTL 477

Query: 452 KKMGLR 457
 K+GLR
 Sbjct: 478 AKLGLR 483

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 1285

A DNA sequence (GBSx1362) was identified in *S.agalactiae* <SEQ ID 3951> which encodes the amino acid sequence <SEQ ID 3952>. This protein is predicted to be shikimate kinase (aroK). Analysis of this protein sequence reveals the following:

55 Possible site: 17
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60

-1421-

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA55181 GB:X78413 shikimate kinase [Lactococcus lactis]
Identities = 65/164 (39%), Positives = 98/164 (59%), Gaps = 8/164 (4%)

5  Query: 1  MPKVLLGFMGVGKTSVANCLENEVIDMDSLIEKHIGMSISRFFTEEGEASFRALQSFLN 60
      M +L+GFMG GK++VA L E D+D LIE+ I M I+ FF GEA FR +E++
      Sbjct: 1  MSIIIGFMGAGKSTVAKLLAEFTDLCLKLIEEEIEMPIATFFELFGEADFRKIENEVFE 60

Query: 61  ELLKKKNEGLVIASGGGIVLLEENRRLTLNRHNNIL-LTGSFEVLYHRIKKDEKNRRPL 119
      ++K ++IA+GGGI+ E + L L+R + ++ LT F+ L+ RI D +N RP
10  Sbjct: 61  LAVQK---DIIIIATGGGII--ENPKNLNVLDRA SRVFLTADFDTLWKRISMDWQNVRP- 114

Query: 120  FLNHSKEEFYDIYQKRMLLYSGLSDMIIDTDYLTPOKIATVIGE 163
      L KE +++KRM YS ++D+ ID +P++IA I E
15  Sbjct: 115 -LAQDKEAAQLLFEKRMKDYSLVADLTIDVTDKSPEQIAEQIRE 157

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3953> which encodes the amino acid sequence <SEQ ID 3954>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
20  >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:CAA55181 GB:X78413 shikimate kinase [Lactococcus lactis]
Identities = 63/160 (39%), Positives = 97/160 (60%), Gaps = 5/160 (3%)

30  Query: 1  MTKVLLGFMGVGKTTVSKHLSMHCKDMDAIIIEAKIGMSIAAFFEQHGEIAFRTIESQVLK 60
      M+ +L+GFMG GK+TV+K L+ D+D +IE +I M IA FFE GE FR IE++V +
      Sbjct: 1  MSIIIGFMGAGKSTVAKLLAEFTDLCLKLIEEEIEMPIATFFELFGEADFRKIENEVFE 60

Query: 61  DLLFANDNSIIVTGGGVVVLQENRQLLRKNHQHNILLVASFETLYQRLKHKKSQRPLFL 120
      L + II TGGG++ +N +L + + L A F+TL++R+ D ++ RP L
35  Sbjct: 61  --LAVQKDIIIIATGGGIIENPKNLNVLDRA SRVFLTADFDTLWKRISMDWQNVRP--L 115

Query: 121  KYSKEAFYEFYQQRMMVFYEGLSDLVIRVDHRTPEEVANII 160
      KEA +++RM Y ++DL I V ++PE++A I
40  Sbjct: 116 AQDKEAAQLLFEKRMKDYSLVADLTIDVTDKSPEQIAEQI 155

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 88/161 (54%), Positives = 120/161 (73%), Gaps = 1/161 (0%)

45  Query: 1  MPKVLLGFMGVGKTSVANCLENEVIDMDSLIEKHIGMSISRFFTEEGEASFRALQSFLN 60
      M KVLGFMGVGKT+V+ L DMD++IE IGMSI+ FF + GE +FR +ESQ L
      Sbjct: 1  MTKVLLGFMGVGKTTVSKHLSMHCKDMDAIIIEAKIGMSIAAFFEQHGEIAFRTIESQVLK 60

Query: 61  ELLKKKNEGLVIASGGGIVLLEENRRLTLNRHNNILLTGSFEVLYHRIKKDEKNRRPLF 120
      +LL N+ +I +GGG+V+L+ENR+LL N +NILL SFE LY R+K D+K++RPLF
50  Sbjct: 61  DLLFA-NDNSIIVTGGGVVVLQENRQLLRKNHQHNILLVASFETLYQRLKHKKSQRPLF 119

Query: 121  LNHSKEEFYDIYQKRMLLYSGLSDMIIDTDYLTPOKIATVI 161
      L +SKE FY+ YQ+RM+ Y GLSD++I D+ TP+++A +I
55  Sbjct: 120 LKYSKEAFYEFYQQRMMVFYEGLSDLVIRVDHRTPEEVANII 160

```

SEQ ID 3952 (GBS152) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 25 (lane 2; MW 20kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 2; MW 45.5kDa).

-1422-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1286

A DNA sequence (GBSx1363) was identified in *Sagalactiae* <SEQ ID 3955> which encodes the amino acid sequence <SEQ ID 3956>. This protein is predicted to be 3-phosphoshikimate 1-carboxyvinyltransferase (aroA). Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.81 Transmembrane 241 - 257 (240 - 257)

INTEGRAL Likelihood = -0.06 Transmembrane 390 - 406 (390 - 406)

----- Final Results -----

bacterial membrane --- Certainty=0.1723(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9673> which encodes amino acid sequence <SEQ ID 9674> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD45819 GB:AF169483 5-enolpyruvylshikimate-3-phosphate synthase

[Streptococcus pneumoniae]

Identities = 288/426 (67%), Positives = 347/426 (80%)

Query: 5 MKLLTNANTLKGTIRVPGDKSISHRAIIFGSISQGVTRIVDVLRGEDVLSTIEAFKQMGV 64

MKL TN L G IRVPGDKSISHR+IIFGS+++G T++ D+LRGEDVLST++ F+ +GV

Sbjct: 1 MKLKTNIRHLGHIIRVPGDKSISHRSIIFGSLAEGETKVYDILRGEDVLSTMQVFRDLGV 60

Query: 65 LIEDDGEIITIYGKGFAGLTQPNLLDMGNSGTSMLRIAGVLAGQEFVMTMGDNLSSKR 124

IED +IT+ G G AGL P N L+MGNSGTS+RLI+GVLAG +FEV M GD+SLSKR

Sbjct: 61 EIEDKDGIVITVQGVGMAGLKAPQNALNMGNSGTSIRLISGVLAGADFEVEMFGDDSLSKR 120

Query: 125 PMDRIALPLSKMGARISGVNTNRLDPLPLKQCTKKLKPIFYHLPVASAQVKSALIFAALQT 184

PMDR+ LPL KMG ISG T RDLPLPL+L+GTK L+PI Y LP+ASAQVKSAL+FAALQ

Sbjct: 121 PMDRVTLPLKKMGVSISGQTERDLPPLRLKGTKNLRPIHYELPIASAQVKSALMFAALQA 180

Query: 185 KGESLIVEKEQTRNHTEDMIRQFGGHLIDKKEIRLNGGQSLVGQDIRVPGDISSAAFVI 244

KGES+I+EKE TRNHTEDM++QFGGHL + K+I + G Q L GQ + VPGDISSAAFV+

Sbjct: 181 KGESVIIIEKEYTRNHTEDMLQQFGGHLSDVGKKITVQGPQKLTGQKVVPVPGDISSAAFV 240

Query: 245 VAGLIIPNSHIIENVGINETRTGILDVVSVMGKIKLSSVDNQVKSATLTVDYSHLQAT 304

VAGLI PNS ++L+NVGINETRTGI+DV+ MGGK++++ +D KSATL V+ S L+ T

Sbjct: 241 VAGLIAPNSRLVLQNVGINETRTGIIDVIRAMGGKLEITEIDPVAKSATLIVESSDLKGT 300

Query: 305 HISGAMIPRLIDELPIIALLATQAQGTTVIADAQELKVKETDRIQVQVVEESLKQMGADITA 364

I GA+IPRLIDELPIIALLATQAQG TVI DA+ELKVKETDRIQVV ++L MGADIT

Sbjct: 301 EICGALIPRLIDELPIIALLATQAQGVTVIKDAEELKVKETDRIQVQVADALNSMGADITP 360

Query: 365 TADGMIIRGNTPLHAASLDCHGDHRIGMMIAIAALLVKEGEVDLSGEEAINTSYNPFLEH 424

TADGMII+G + LH A ++ GDHRIGMM AIAALLV +GEV+L EAINTSYP+F +

Sbjct: 361 TADGMIIRGKSALHGARVNTFGDHRIGMMIAIAALLVADGEVELDRAEAINTSYPSFFDD 420

Query: 425 LEGLVN 430

LE L++

Sbjct: 421 LESLIH 426

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3957> which encodes the amino acid sequence <SEQ ID 3958>. Analysis of this protein sequence reveals the following:

Possible site: 36

-1423-

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.18 Transmembrane 240 - 256 (239 - 256)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAD45819 GB:AF169483 5-enolpyruvylshikimate-3-phosphate synthase
 [Streptococcus pneumoniae]
 Identities = 278/426 (65%), Positives = 346/426 (80%)

15 Query: 4 MKLRTNAGPLQGTIQVPGDKSISHRAVILGAVAKGETRVKGLLKGEDVLSTIQAFRNIGV 63
 MKL+TN L G I+VPGDKSISHR++I G++A+GET+V +L+GEDVLST+Q FR+LGV
 Sbjct: 1 MKLKTNIIRHLHGIIIRVPGDKSISHRSIIIFGSLAEGETKVYDILRGEDVLSTMQVFRDLGV 60

20 Query: 64 RIEEKDDQLVIEGQGFQGLNAPCQTLNMGNSGTSMRLLIAGLAGQPFVSKMIGDESLSKR 123
 IE+KD + ++G G GL AP LNMGNSGTS+RLI+G+LAG F V+M GD+SLSKR
 Sbjct: 61 EIEDKDGVITVQGVGMAGLKAPQNALNMGNSGTSIRLISGVLGADFEVEMFGDDSLSKR 120

25 Query: 124 PMDRIVYPLKQMGVEISGETDRQFPPLQLQGNRNLPITYTLPISSAQVKSAILLAALQA 183
 PMDR+ PLK+MGV ISG+T+R PPL+L+G +NL+PI Y LPI+SAQVKSAAA AALQA
 Sbjct: 121 PMDRVTLPLKKMGVSIISGQTERDLPLRLKGTKNLRPIHYELPIASAQVKSALMFAALQA 180

30 Query: 184 KGTTQVVEKRIITRNHTEEMIQQFGGRLIVDGKRITLVGPQQLTAQEITVPGDISSAAFVWL 243
 KG + ++EKE TRNHTE+M+QQFGG L VDGK+IT+ GPQ+LT Q++ VPGDISSAAFVWL
 Sbjct: 181 KGESVIEKEYTRNHTEMDLQQFGGHLSDVGKKITVQGPQKLTGQKVVVPGDISSAAFVWL 240

35 Query: 244 VAGLIIPGSELLKKNVGNPRTTGILEVVEKMGQAQIVYEDMNKKEQVTSIRVVYSNMKGT 303
 VAGLI P S L+L+NVG+N TRTGI++V+ MG ++ +++ + ++ V S++KGT
 Sbjct: 241 VAGLIAPNSRLVLQNVGINETRTGIIIDVIRAMGGKLEITEIDPVAKSATLIVESSDLKGT 300

40 Query: 304 IISGGLIPRLIDELPIIALLATQAQGTTCIKDAQELRVKETDRIQVVTDIINSMGANIKA 363
 I G LIPRLIDELPIIALLATQAQ T IKDA+EL+VKETDRIQV D LNSMGA+I
 Sbjct: 301 EICGALIPRLIDELPIIALLATQAQGVTVIKDAEELVKETDRIQVVDALNSMGADITP 360

45 Query: 364 TADGMIKGPITVLYGANTSTYGDHRIGMMTAIAALLVKQGVHLDKBEAINTSYPTFFKD 423
 TADGMIK + L+GA +T+GDHRIGMMTAIAALLV G+V LD+ EAI TSYP+FF D
 Sbjct: 361 TADGMIKGSALHGARVNTFGDHRIGMMTAIAALLVADGEVELDRAEAINTSYPSFFDD 420

Query: 424 LERLCH 429
 LE L H
 Sbjct: 421 LESLIH 426

An alignment of the GAS and GBS proteins is shown below.

Identities = 269/424 (63%), Positives = 331/424 (77%)

50 Query: 5 MKLLTNANTLKGITIRVPGDKSISHRAIIFGSIQGVTRIVDVLRGEDVLSTIEAFKQMGV 64
 MKL TNA L+GTI+VPGDKSISHRA+I G++++G TR+ +L+GEDVLSTI+AF+ +GV
 Sbjct: 4 MKLRTNAGPLQGTIQVPGDKSISHRAVILGAVAKGETRVKGLLKGEDVLSTIQAFRNIGV 63

55 Query: 65 LIEDDGEIITIYGKGFAGLTQPNLLDMGNSGTSMRLLIAGVLAGQEFVMTVGDNSLSKR 124
 IE+ + + I G+GF GL P L+MGNSGTSMRLLIAG+LAGQ F V M+GD SLSKR
 Sbjct: 64 RIEEKDDQLVIEGQGFQGLNAPCQTLNMGNSGTSMRLLIAGLAGQPFVSKMIGDESLSKR 123

60 Query: 125 PMDRIALPLSKMGARISGVTNRDLPLKLQGTKKLKPIFYHLEVASAQVKSALIFAALQT 184
 PMDRI PL +MG ISG T+R PPL+LQ + L+PI Y LP++SAQVKSAAA AALQ
 Sbjct: 124 PMDRVYPLKQMGVEISGETDRQFPPLQLQGNRNLPITYTLPISSAQVKSAILLAALQA 183

65 Query: 185 KGESLIVEKEQTNRNHTEDMIRQFGGHLDIKDKEIRLNGGQSLVGQDIRVPGDISSAAFVWL 244
 KG + +VEKE TRNHTE+MI+QFGG L + K I L G Q L Q+I VPGDISSAAFVWL
 Sbjct: 184 KGTTQVVEKRIITRNHTEEMIQQFGGRLIVDGKRITLVGPQQLTAQEITVPGDISSAAFVWL 243

Query: 245 VAGLIIPNSHIILENVGINETRTGILDVVSMMGGKIKLSSVDNQVKSATLTVDYSHLQAT 304
 VAGLIIP S ++L+NVG+N TRTGIL+VV KMG +I ++ + + ++ V YS+++ T

-1424-

Sbjct: 244 VAGLIIPGSELLKKNVGVNPTRTGILEVVEKMGAIQVYEDMNKKEQVTSIRVVYSNMKGT 303

Query: 305 HISGAMIPRLIDELPIALLATQAQGTTVIADAQELKVKETDRIQVVVESLKQMGADITA 364
ISG +IPRLIDELPIALLATQAQGT I DAQEL+VKETDRIQVV + L MGA+I A

5 Sbjct: 304 IISGGLIPRLIDELPIALLATQAQGTTCIKDAQELRVKETDRIQVVTDILNSMGANIK 363

Query: 365 TADGMIIRGNTPLHAASLDCHGDRIGMMIAAALLVKEGEVDLSGEEAINTSYPNFLEH 424
TADGMII+G T L+ A+ +GDHRIGMM AIAALLVK+G+V L EEA TSYP F +

10 Sbjct: 364 TADGMIIGPTVLYGANTSTYGDHRIGMMTAIAALLVKQGQVHLDKEEAIMTSYPTFFKD 423

Query: 425 LEGL 428
LE L

Sbjct: 424 LERL 427

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1287

A DNA sequence (GBSx1364) was identified in *S. agalactiae* <SEQ ID 3959> which encodes the amino acid sequence <SEQ ID 3960>. Analysis of this protein sequence reveals the following:

20 Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -1.12 Transmembrane 6 - 22 (6 - 22)

25 ----- Final Results -----
bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:AAF20148 GB:AF208390 actinin-like protein [Entamoeba histolytica]
Identities = 62/236 (26%), Positives = 107/236 (45%), Gaps = 38/236 (16%)

35 Query: 144 NYNSTNSSNPESMLFYEQQLKTWLSSTH---KNYYLDYK--VTPIYQNNELIPRKIELK- 196
N N + N + + L W+++ N+ D+K V + + +I+ +
Sbjct: 116 NANQKKNVNAKEEVVENNALLDWNSFGLNVSNFSSDWKDGVALVKLTEAVSAGQIKFEQ 175

40 Query: 197 YVGIDKTGKLLPIFIGNKSTQDQFGI-----STVTLENTSPNATIDYLSGKAQN----- 245
+ G+D T ++ K +QF I + E P + + Y+S +
Sbjct: 176 FSGLDNTQMVIDC---QKLAYEQFKIPIILMDVKDLVCERPDPKSIIMTYVSVYKERYEQLL 232

45 Query: 246 TVLSAKEQRKLIKHEEEKRLAEK-----KVEEEKAAAETQKKL-EEEQARLAAEAQ-RK 298
KE+++ IA+ E+E++ E+ + E+E+ A E Q++L EEQ RLA E Q RK
Sbjct: 233 VEKEQKEEQERIAREEQERKQKEEQERLAREEQERLAREEQERLAREEQERLAREEQERK 292

50 Query: 299 QKEEQARLAAETQKKQETLVQEQTSQGYKRDYRGRWHRPNGQYASKAEIAAAGLQW 354
QKEEQ RLA E Q++++ QE+ +Q +P Q + + AA W
Sbjct: 293 QKEEQERLAREEQERKQREEQERLNQ-----QOPTSQQLTFFSVQAAADAW 338

50 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3961> which encodes the amino acid sequence <SEQ ID 3962>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have a cleavable N-term signal seq.

55 ----- Final Results -----
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60

-1425-

The protein has homology with the following sequences in the databases:

```

>GP:CAA03161 GB:A49208 unnamed protein product [Streptococcus
pyogenes]
Identities = 54/222 (24%), Positives = 93/222 (41%), Gaps = 39/222 (17%)
5
Query: 44 HYKNTVSSKLLP--FTANYQLQLGELDNLNRA-----TFSHIQLQDRHETKDVRTKINYD 96
+YK +S++ P F + +LD L R T ++ ++ + + K N +
Sbjct: 76 YYKTLGTSQITPALFPKAGDILYSKLDDELGRTRTARGTLTYANVEGSYGVRQSFGK-NQN 134

10
Query: 97 PVGWHN-----YQFPYGDG-SKSSWVMNRGHLVGYQFCGLNDEPRNLVAMTAWLNTGAY 149
P GW Y+ + +G S NR HL+ G + + + A T
Sbjct: 135 PAGWTGNPNHVKYKIEWLNGLSYVGDFWNRSHLIADSLGG-----DALRVNAVGTGTRTQ 188

15
Query: 150 SGANDSNPEGMLYYENRLDSWLALHPDFWLDYKVTPIYSGNEVVPRQIELQYVVGIDSSGE 209
+ GM Y E R WL + D +L Y+V PIY+ +E++PR +
Sbjct: 189 NVGGRDQKGMRYTEQRAQEWEANRDGYLYYEVAPIYNADELIPRAV----- 236

Query: 210 LLTIRLNSNKESIDENGVTITVILENSAPNINLDYLNGTATPK 251
+ + S+ +I+E V++ N+A ++Y NGT T K
20
Sbjct: 237 --VSMQSSDNTINEK----VLVYNTANGYTINYHNGTPTQK 272

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 117/245 (47%), Positives = 166/245 (67%), Gaps = 4/245 (1%)
25
Query: 2 KRKQFIKLGIAITLLTVISLYTPINLATNHTTENIVTAQEY--KTKENGTLPFKHKRQLVL 59
K+K + + LL++ ++ A T N+ A + T + LPF QL L
Sbjct: 5 KQKASLLTAVLILLLSLSTITTITVDAARVRTYPNVSHANTHYKNVTSSKLLPFTANYQLQL 64

30
Query: 60 GELDDKGRATFAHIQLKVKDEPKKRVKRLKTPVGVWHNFKFPYNDGTQKAWLMSRGLI 119
GELD+ RATF+HIQL+ + E K R K + PVGWHN++F Y DG++ +W+M+RG L+
Sbjct: 65 GELDNLNLRATFSHIQLQDRHETKDVRTK-INYDVPVGVWHNYQFPYGDGSKSSWVMNRGHLV 123

Query: 120 CHQFSGLNNERKNLVPMINWLNTGNYNSTNSSNPESMLFYEQKLTWLSTHKNYYLDYKV 179
+QF GLN+E +NLV MT WLNTG Y+ N SNPE ML+YE +L +WL+ H +++LDYKV
35
Sbjct: 124 GYQFCGLNDEPRNLVAMTAWLNTGAYSGANDSNPEGMLYYENRLDSWLALHPDFWLDYKV 183

Query: 180 TPIYQNNELIPRKIELKYVGIDKTGKLLPIFI-GNKSTQDQFGISTVTLENTS PNATIDY 238
TPIY NE++PR+IEL+YVGID +G+LL I + NK + D+ G++TV LEN++PN +DY
40
Sbjct: 184 TPIYSGNEVVPRQIELQYVVGIDSSGELLTIRLNSNKESIDENGVTITVILENSAPNINLDY 243

Query: 239 LSGKA 243
L+G A
Sbjct: 244 LNGTA 248

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 7263> which encodes amino acid sequence <SEQ ID 7264>. An alignment of the GAS and GBS sequences follows:

```

Score = 58.9 bits (140), Expect = 2e-11
Identities = 34/103 (33%), Positives = 55/103 (53%), Gaps = 1/103 (0%)
50
Query: 1 MPFKTNLKAGILLYAMFMASIFLLVLQVYLSQVTALHKEYQAQTDYVKARLIAEIVYQD- 59
M K LKAGILL A+ +A++F LVLQ YL+++ A ++Y +Q + KA L A++ Y+
Sbjct: 1 MILKKKLKAGILLQAIVLAAVFTLVLFQFYLARILATERQYHSQIEASKAYLTAQLAYKTI 60

Query: 60 HRYKASNPVFFKGGQVICRERKERWMLIVKLDQQRQYQFEYLYK 102
S +F GG + + V LD+ Y ++ +
55
Sbjct: 61 EGDSTSGKCYFTGGYASYLQEGNYLQVKVTLDDKGGNYNHKFYR 103

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1426-

Example 1288

A DNA sequence (GBSx1365) was identified in *S.agalactiae* <SEQ ID 3963> which encodes the amino acid sequence <SEQ ID 3964>. This protein is predicted to be enolase (eno). Analysis of this protein sequence reveals the following:

```

5   Possible site: 43
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3025(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:BAA81815 GB:AB029313 enolase [Streptococcus intermedius]
    Identities = 396/435 (91%), Positives = 414/435 (95%), Gaps = 1/435 (0%)

    Query: 1   MSIIITDVYAREVLDSEGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYG 60
              MSIIITDVYAREVLDSEGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYG
20   Sbjct: 1   MSIIITDVYAREVLDSEGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYG 60

    Query: 61   GLGTQKAVDNVNNVIAEAIIGYDVRDQQAIDRAMIALDGTTPNKGKLGANAILGVSIAR 120
              GLGTQKAVDNVNN+IAEA+IGYDVRDQQAIDRAMIALDGTTPNKGKLGANAILGVSIAR
25   Sbjct: 61   GLGTQKAVDNVNNIIAEAVIGYDVRDQQAIDRAMIALDGTTPNKGKLGANAILGVSIAR 120

    Query: 121  AAADYLEVPLYSYLGGFNTKVLPTPMMNIINGGSHSDAPIAFQEFMIMPVGAPTFKEALR 180
              AAADYLE+PLYSYLGGFNTKVLPTPMMNIINGGSHSDAPIAFQEFMI+P GAPTFKEALR
30   Sbjct: 121 AAADYLEIPLYSYLGGFNTKVLPTPMMNIINGGSHSDAPIAFQEFMIVPAGAPTFKEALR 180

    Query: 181  WGAEVFHALKKILKRGLETAVGDEGGFAPKFEGTEDGVETILKAIEAAGYFAGENGIMI 240
              WGAE+PHALKKILK RGL TAVGDEGGFAP+F+GTEDGVETIL AIEAAGY G++ + +
35   Sbjct: 181 WGAEIFHALKKILKSRGLATAVGDEGGFAPRFDGTEDGVETILAAIEAAGYVPGKD-VFL 239

    Query: 241  GFDCASSEFYDAERKVYDYTSKFEGEGGAVRTAAEQIDYLEELVNKYPIITIEDGMDENDW 300
              GFDCASSEFYD ERKVYDY+KFEGEG AVRTA EQIDYLEELVNKYPIITIEDGMDENDW
40   Sbjct: 240 GFDCASSEFYDKERKVYDYTSKFEGEGAAVRTADEQIDYLEELVNKYPIITIEDGMDENDW 299

    Query: 301  DGWKALTERLGGRVQLVGDDFFVTNTDYLARGIKEEAANSILIKVNQIGTLTETFEAIE 360
              DGWK LTERLG +VQ VGDDFFVTNT YL +GI E ANSILIKVNQIGTLTETF+AIE 360
45   Sbjct: 300 DGWKKLTERLGKQVQVGGDDFFVTNTSYLEKGINEACANSILIKVNQIGTLTETFDAIE 359

    Query: 361  AKEAGYTAVVSHRSGETEDSTIADIATNAGQIKTGSLSRTDRIAKYNQLLRIBDQLGE 420
              AKEAGYTAVVSHRSGETEDSTIADIATNAGQIKTGSLSRTDRIAKYNQLLRIBDQLGE
50   Sbjct: 360 AKEAGYTAVVSHRSGETEDSTIADIATNAGQIKTGSLSRTDRIAKYNQLLRIBDQLGE 419

    Query: 421  VAQYKGIKSFYNLKK 435
              VA+Y+G+KSFYNL K
60   Sbjct: 420 VAEYRGLKSFYNLSK 434

```

Proteins in the glycolysis/gluconeogenesis pathway have been experimentally detected on the surface of *Streptococci*.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3965> which encodes the amino acid sequence <SEQ ID 3966>. Analysis of this protein sequence reveals the following:

```

    Possible site: 43

55   >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.3025(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
60   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```


The protein has homology with the following sequences in the databases:

>GP:BAA81815 GB:AB029313 enolase [Streptococcus intermedius]
Identities = 396/435 (91%), Positives = 415/435 (95%), Gaps = 1/435 (0%)

5 Query: 1 MSIIITDVYAREVLD SRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYL 60
MSIIITDVYAREVLD SRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRY
Sbjct: 1 MSIIITDVYAREVLD SRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYG 60

10 Query: 61 GLGTQKAVDNVNNIIAEAIIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR 120
GLGTQKAVDNVNNIIAEAIIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR
Sbjct: 61 GLGTQKAVDNVNNIIAEAVIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR 120

15 Query: 121 AAADYLEVPLTYLGGFNTKVLPTPMNI INGGSHSDAPIAQEFMIMPVGAPT FKEGLR 180
AAADYLE+PLY+YLGGFNTKVLPTPMNI INGGSHSDAPIAQEFMI+P GAPTFKE LR
Sbjct: 121 AAADYLEIPLYSYLGGFNTKVLPTPMNI INGGSHSDAPIAQEFMIVPAGAPT FKEALR 180

20 Query: 181 WGAEVFHALKKILKERGLVTAVGDEGGFAPKFEGTEDGVETILKAIEAAGYEAGENGIMI 240
WGAE+FHALKKILK RGL TAVGDEGGFAP+F+GTEDGVETIL AIEAAGY G++ + +
Sbjct: 181 WGAEIFHALKKILKSRGLATAVGDEGGFAPRFDGTEDGVETILAAIEAAGYVPGKD-VFL 239

25 Query: 241 GFDCASSEFYDKERKVYDYTKFE GEGA AVRTSAEQVDYLEELVNKYPIITIEDGMDENDW 300
GFDCASSEFYDKERKVYDYTKFE GEGA AVRT+ EQ+DYLEELVNKYPIITIEDGMDENDW
Sbjct: 240 GFDCASSEFYDKERKVYDYTKFE GEGA AVRTADEQIDYLEELVNKYPIITIEDGMDENDW 299

30 Query: 301 DGWKVLTERLGKRVQLVGDDFFVTNT EYLARGIKENAANSILIKVNQIGTLTETFEA IEM 360
DGWK LTERLGK+VQ VGDDFFVTNT YL +GI E ANSILIKVNQIGTLTETTF+A IEM
Sbjct: 300 DGWKVLTERLGKRVQLVGDDFFVTNTSYLEKGINEACANSILIKVNQIGTLTETTFDA IEM 359

35 Query: 361 AKEAGYTAVVSHRSGETEDSTIADI AVATNAGQIKTGSLSRTDRIAKYNQLLR IEDQLGE 420
AKEAGYTAVVSHRSGETEDSTIADI AVA NAGQIKTGSLSRTDRIAKYNQLLR IEDQLGE
Sbjct: 360 AKEAGYTAVVSHRSGETEDSTIADI AVAANAGQIKTGSLSRTDRIAKYNQLLR IEDQLGE 419

Query: 421 VAQYKGIKSFYNLKK 435
VA+Y+G+KSFYNL K
Sbjct: 420 VAEYRGLKSFYNLSK 434

An alignment of the GAS and GBS proteins is shown below.

Identities = 421/435 (96%), Positives = 427/435 (97%)

40 Query: 1 MSIIITDVYAREVLD SRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYG 60
MSIIITDVYAREVLD SRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRY
Sbjct: 1 MSIIITDVYAREVLD SRGNPTLEVEVYTESGAFGRGMVPSGASTGEHEAVELRDGDKSRYL 60

45 Query: 61 GLGTQKAVDNVNNVIAEAIIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR 120
GLGTQKAVDNVNN+IAEAIIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR
Sbjct: 61 GLGTQKAVDNVNNIIAEAIIGYDVRDQQAIDRAMIALDGT PNKGKLGANAILGVSI AVAR 120

50 Query: 121 AAADYLEVPLYSYLGGFNTKVLPTPMNI INGGSHSDAPIAQEFMIMPVGAPT FKEALR 180
AAADYLEVPLY+YLGGFNTKVLPTPMNI INGGSHSDAPIAQEFMIMPVGAPT FKE LR
Sbjct: 121 AAADYLEVPLTYLGGFNTKVLPTPMNI INGGSHSDAPIAQEFMIMPVGAPT FKEGLR 180

55 Query: 181 WGAEVFHALKKILKERGLETAVGDEGGFAPKFEGTEDGVETILKAIEAAGYEAGENGIMI 240
WGAEVFHALKKILKERGL TAVGDEGGFAPKFEGTEDGVETILKAIEAAGYEAGENGIMI
Sbjct: 181 WGAEVFHALKKILKERGLVTAVGDEGGFAPKFEGTEDGVETILKAIEAAGYEAGENGIMI 240

60 Query: 241 GFDCASSEFYDAERKVYDYSKFE GEGG AVRTAAEQIDYLEELVNKYPIITIEDGMDENDW 300
GFDCASSEFYD ERKVYDY+KFE GEG AVRT+AEQ+DYLEELVNKYPIITIEDGMDENDW
Sbjct: 241 GFDCASSEFYDKERKVYDYTKFE GEGA AVRTSAEQVDYLEELVNKYPIITIEDGMDENDW 300

65 Query: 301 DGWKALTERLGGRVQLVGDDFFVTNTDY LARGIKEEAANSILIKVNQIGTLTETFEA IEM 360
DGWK LTERLG RVQLVGDDFFVTNT+YLARGIKE AANSILIKVNQIGTLTETFEA IEM
Sbjct: 301 DGWKVLTERLGKRVQLVGDDFFVTNT EYLARGIKENAANSILIKVNQIGTLTETFEA IEM 360

Query: 361 AKEAGYTAVVSHRSGETEDSTIADI AVATNAGQIKTGSLSRTDRIAKYNQLLR IEDQLGE 420
AKEAGYTAVVSHRSGETEDSTIADI AVATNAGQIKTGSLSRTDRIAKYNQLLR IEDQLGE

-1428-

Sbjct: 361 AKEAGYTAVVSHRSGETEDSTIADIATNAGQIKTGSLSRDRIAKYNQLLRIEDQLGE 420

Query: 421 VAQYKGIKSFYNLKK 435

VAQYKGIKSFYNLKK

5 Sbjct: 421 VAQYKGIKSFYNLKK 435

SEQ ID 3964 (GBS311) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 3; MW 51kDa).

GBS311-His was purified as shown in Figure 203, lane 10.

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1289

A DNA sequence (GBSx1366) was identified in *S.agalactiae* <SEQ ID 3967> which encodes the amino acid sequence <SEQ ID 3968>. Analysis of this protein sequence reveals the following:

15 Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.1998(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1290

A DNA sequence (GBSx1367) was identified in *S.agalactiae* <SEQ ID 3969> which encodes the amino acid sequence <SEQ ID 3970>. This protein is predicted to be di-/tripeptide transporter. Analysis of this protein sequence reveals the following:

30 Possible site: 54
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -14.33	Transmembrane	93 - 109 (87 - 122)
INTEGRAL	Likelihood = -9.02	Transmembrane	117 - 133 (110 - 141)
35 INTEGRAL	Likelihood = -8.44	Transmembrane	333 - 349 (328 - 353)
INTEGRAL	Likelihood = -5.84	Transmembrane	19 - 35 (17 - 38)
INTEGRAL	Likelihood = -3.08	Transmembrane	151 - 167 (151 - 167)
INTEGRAL	Likelihood = -2.55	Transmembrane	264 - 280 (264 - 281)
INTEGRAL	Likelihood = -2.28	Transmembrane	44 - 60 (44 - 60)
40 INTEGRAL	Likelihood = -2.02	Transmembrane	238 - 254 (238 - 255)

----- Final Results -----

45 bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9395> which encodes amino acid sequence <SEQ ID 9396> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1429-

>GP:CAB12175 GB:Z99106 similar to di-tripeptide ABC transporter
(membrane protein) [Bacillus subtilis]
Identities = 175/359 (48%), Positives = 254/359 (70%), Gaps = 9/359 (2%)

5 Query: 1 MVGNLYGENDSRRDAGFSIFVFGINLGAFISPIVVGYLQGEVNFHLGFSLAAGMFFGLL 60
+VG+LY + D RRD+GFSIF GINLG ++P++VG LGQ+ N+HLGF AA+GM GL+
Sbjct: 142 VVGDLTYTKEDPRRDSGFSIFYMGINLGGLLAPLIVGTGQKYNHHLGFGAAVGMMLGLI 201

10 Query: 61 QYTLDGKKYLTEESLRPNPLSPEEKSSLYKKVGLILIGIVIVLILLHLMHMLTIEVIID 120
+ L KK L +PLS +KS++ +G+I++ I +++ + +LTI+ ID
Sbjct: 202 VFPLTRKKNLGLAGSNVNPPLS--KKSAGTGTGIVIIVAIAVVISVQ--TGVLTIKRFID 257

15 Query: 121 IPSIIAIAIPIIYFIKILSSKKISSVERSRVWAYIPLFIASILFWSIERQGSVVLALFAD 180
+ SI+ I IP+IYFI + +SKK E+SR+ AY+PLFT +++FW+I+EQG+ +LA++AD
Sbjct: 258 LVSILGILIPVIYFIIMFTSKKADKTEKSRLAAYVPLFIGAVMFVAIQEQGATILAVYAD 317

20 Query: 181 EQTKLYLNFFGHHINFPSSYFQSMNPLFIMLYVPFFAWLWAKWGSQPSPPKKFAYGLFF 240
E+ +L L F SS+FQS+NPLF++++ P FAWLW K G +QPS+P KF+ G+
Sbjct: 318 ERIRLSLGGF----ELQSSWFQSLNPLFVVI FAPIFAWLWMKLGKRQPSPTPVKFSIGIIL 373

25 Query: 241 AGASFLWMLPGLLFGVNAKVSPLWLTMSWAIVIVGEMLISPVGLSATSKLAPKAFOAQM 300
AG SF+ M+ P + G A VSPLWL +S+ +V++GE+ +SPVGLS T+KLAP AF AQ
Sbjct: 374 AGLSPTIMVFPAMQ-GKEALVSPLWLVLVLSFLLVVLGELCLSPVGLSVTTKLAPAFSAQT 432

30 Query: 301 MSIWFLSNAAAQAINAQIVKLYTPDTOTLYYGVVGGITVVFILLFYVPRIEKLMSGV 359
MS+WFL+NAAAQAINAQ+ L+ +T+Y+G +G I++V G ILL P I++ M GV
Sbjct: 433 MSMWFLTNAAAQAINAQVAGLFDKIPETMYFGTIGLISIVLGGIILLSPVIKRAMKGV 491

No corresponding DNA sequence was identified in *S.pyogenes*.

- 30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1291

A DNA sequence (GBSx1369) was identified in *S.agalactiae* <SEQ ID 3971> which encodes the amino acid sequence <SEQ ID 3972>. Analysis of this protein sequence reveals the following:

35 Possible site: 37
>>> Seems to have an uncleavable N-term signal seq

40 ----- Final Results -----
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1292

A DNA sequence (GBSx1370) was identified in *S.agalactiae* <SEQ ID 3973> which encodes the amino acid sequence <SEQ ID 3974>. Analysis of this protein sequence reveals the following:

50 Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2485(Affirmative) < succ>

-1430-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAF61315 GB:U96166 unknown [Streptococcus cristatus]
 Identities = 181/442 (40%), Positives = 270/442 (60%), Gaps = 2/442 (0%)

Query: 1 MINLFDSTYTQSSWDLHFSLIKSGYINPTIALNDDGFLPDDVTSPYLYTGFAGTGRPL 60
 MI LFD Y Q+S+DL SL +G P + + DDG+L DV SPY Y+TG T GRP+

10 Sbjct: 1 MICLFDRTYQASFDLLRSLKATGLDCPVVVQDDGYLSPDVESPYSYFTGDLDTPEGRPI 60

Query: 61 YYNELRVPTWEIIGFSSGADIVDLGVKKGRITIANPNHKRLIKEVDWDFEQGRVILKDR 120
 Y+N + P WEI + +I+D+G K+ I Y P H+R ++ V+W D +G+V D

15 Sbjct: 61 YFNLVPHLWEIRSSNVNGEILDMGKKRANIFYRQPTHERVRVAVEWLDTGQVRAADI 120

Query: 121 FNKFGFCFAQTTFYNADGQAIQTSYYNKDRQEVISENHMTGDYILNDNNQFKVFKSKVEFV 180
 +N+ G FAQ Y+ + T Y+++ VI ENH+TGD IL + +FKSK EKV

20 Sbjct: 121 YNRKGRLEAQITYDQTRPTHTRYFDQSNVVMENHILTGDIILTLGKRIHIFKSKQEFV 180

Query: 181 INYLQEAQFNLDRIFYNSLSTPFLVSYL--NRLESKDVLFWQEPVDDIPGNMRLLLNN 238
 + YLQ ++ DRI YNSL+TPFLV++ L ++DVLFWQEP+ + +PGNM++ +

25 Sbjct: 181 VFYLQYRGYDTRIIYNSLATPFLVAYALRPKNGRAEDVLFWQEPGEALPGNMKVAMKM 240

Query: 239 PSPNTKIVIQSYEAYANAMRLLTDEEQKQVSFLGFMYPKIKETEKLNQALILTNSDQIEA 298
 P N +I +Q + Y L T EE+ +G++Y + ++ +ALILTNSDQ+E

30 Sbjct: 241 PHNRIRIAVQDRQVYEKIQSLATPEEKVYFHNIGYIYDYQRLNNMNPEALILTNSDQLEQ 300

Query: 299 LESLVTSLPNTFFNIGALTEMSSDLNMFNGKYDNVVLYPNITTNQIQYLSNICAFYLDINH 358
 +E L+T LPN+ F+IGA+TEMS LM +Y NV LYPNI ++ L C YLDIN

35 Sbjct: 301 IEQLLTQLPNVHFHIGAITEMSGHLMGLNRYPNVSLYPNIRPAKVAELFERCDLYLDINI 360

Query: 359 HNEILSAVRSFAFEHQQLIFAFEETSHQIRFVSPKNIFFPKDIFTFISHLQPLIGNKCNE 418
 +EIL+A R+AFE+ LI +F T H RF++ +I+ +++ + +Q + + +E

40 Sbjct: 361 SDEILNACRTAFENMLILSFTNTCHSRRFIADDDHIYAPENVSGMVDKIQSALAHSSSEME 420

Query: 419 KALKQQLEDCHVSSSTQYQSVI 440
 AL +Q + + +S QY+++I

Sbjct: 421 AALTRQQAANQASLEQYKAI 442

40 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1293

45 A DNA sequence (GBSx1371) was identified in *S.agalactiae* <SEQ ID 3975> which encodes the amino acid sequence <SEQ ID 3976>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.06 Transmembrane 405 - 421 (404 - 422)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA94320 GB:AB033763 hypothetical protein [Staphylococcus aureus]
 Identities = 66/195 (33%), Positives = 99/195 (49%), Gaps = 9/195 (4%)

-1431-

Query: 259 NYYDYQFTNANRDFDFTTSTDKQTELEQQFKQFTNHNPRIITIPVGSID----NLKMMP 314
 N Y + F N NR+ I ST +Q + N+ + TIPVG ID NLK
 Sbjct: 15 NTYKHVFNLNRYSGIIVSTKQ---QLDISARINNEIPVHTIPVGIDEHFTNLKRNN 70

5 Query: 315 DNRRPYSILTASRLASEKHVDWLVRVIRIREILPEVTFDIYSGSGEEEEKIRNIINAANA 374
 + I++ +R + EK ++ + V ++ + P + +YG G EEEK + +I N
 Sbjct: 71 HSINNKKIISVARYSPEKQLNHQIELVSKLIKEFPNIRLHLYGFGKEEEKYKQLITEYNL 130

10 Query: 375 TEYIRLMG-HKNLSNVYQNYELYLTASKSEGFLTLLEAIGAGLPLIGFDVRYGNQTFIK 433
 + L G +NLS Q+ + L S EGF L LLE I G+P +G++ +YG I
 Sbjct: 131 ENNVFLRGFRRNLSAEIQDAYMSLITSNMEGFNLGLLETITTEGIPFVGYNISKYGPSELIL 190

Query: 434 DGENGYLIPRFDMD 448
 + ENGYLI + D D+
 15 Sbjct: 191 NNENGYLINKNDKDE 205

SEQ ID 3976 (GBS426) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 4; MW 58.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 3; MW 84kDa).

20 GBS426-GST was purified as shown in Figure 220, lane 5.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1294

A DNA sequence (GBSx1372) was identified in *S.agalactiae* <SEQ ID 3977> which encodes the amino acid sequence <SEQ ID 3978>. This protein is predicted to be preprotein translocase *seca* subunit (*secA*). Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.69 Transmembrane 75 - 91 (75 - 91)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44957 GB:U56901 involved in protein export [Bacillus subtilis]
 Identities = 336/794 (42%), Positives = 506/794 (63%), Gaps = 29/794 (3%)

40 Query: 5 NSLFSLDKKRLKKLQRTLTNTINSLKGQMATLSNEELQAKTTEFRKRLVNGETLDDICAEA 64
 N +F K+ L + ++ N I++++G LS++ I+ KT EF++RL G T DD+ EA
 Sbjct: 6 NKMPDPTKRRLNRYEKIANDIDATRGDYENLSDALKHKKTIEFKERLEKGATTDLLVEA 65

45 Query: 65 FAVVREADERVLGLFPYDVQVIGGLVLHQGNTAEMKTGEGKTLTATMPLYLNALEGKGAM 124
 FAVVREA RV G+FP+ VQ++GG+ LH GN AEMKTGEGKTLT+T+P+YLNAL GK
 Sbjct: 66 FAVVREASRRVTGMFPFKVQLMGGVALHDGNIAEMKTGEGKTLTSTLPVYLNALTGKGVH 125

50 Query: 125 LLTNNSYLAIRDAEEMGVYRFLGLSVGVGVSDNEBEDRDAATKRAVYSSDIVYSTSSAL 184
 ++T N YLA RDAE+MGK++ FLGL+VG+ ++ +++ KR Y++DI YST++ L
 Sbjct: 126 VVTVNEYLASRDAEQMGKIFEFLGLTVGLNLSMSKDE-----KREAYAADITYSTNNEL 180

55 Query: 185 GFDYLIDNLASSKSQKYPKLHYAIVDEADAVLLDMAQTPLVISGSPRVQSNLYKIADEL 244
 GFDYL DN+ K Q LH+A++DE D++L+D A+TPL+ISG + LY A+
 Sbjct: 181 GFDYLRDNMVLYKEQMVQRPLHFAVIDEVDLSILIDEARTPLIISGQAAKSTKLYVQANAF 240

Query: 245 ILSFEEQVDYDFDKERQEVWIKNQGVREABRYFRIPHFKQSNRELVRHLNLSLKAHKLF 304
 + + + + DY +D + + V + +G+ +AE+ F I + + + L H+N +LKAH
 Sbjct: 241 VRTLKAEKDYYDIKTKAVQLTEGMTKAEKAFGIDNLFVDVKHVALNHHINQALKAHVAM 300

-1432-

Query: 305 ERGKDYVVDGEIKLLDATNGRVLEGTKLQGGVHQAEQKEHLNVTPESTRAMASITYQNL 364
 ++ DYVV+DG++ ++D+ GR+++G + G+HQAE KE L + ES +A+IT+QN
 Sbjct: 301 QKDVDYVVEDGQVVIVDSFTGRMLMKGRYSEGLHQAEAKEGLEIQNESMTLATITTFQNY 360

Query: 365 FRMFTKLAGMTGTGKTAEKEFIEVDMEVVRIPITNSPVRRIDYDPDKIYTTLPKIHATIE 424
 FRM+ KLAGMTGT KT E+EF +Y+M+VV IPTN PV R D PD IY T+ K A E
 Sbjct: 361 FRMYEKLAGMTGTAKTEEBEFRIYNMQVVTIPTNRPVVRDDRPDLIYRTMEGKFKAVAE 420

Query: 425 FVKQVHDTGQPILLVAGSVRMSSELSLLSGIPHSLLNAQSAVKEAQMAEAGQKGAV 484
 V Q + TGQP+L+ +V SEL S+LL GIPH +LNA++ +BAQ+I EAGQKGAV
 Sbjct: 421 DVAQRYMTGQPVLVGTAVETSELISKLLKNKGIPHQVLNAKNHREBAQITEEAGQKGAV 480

Query: 485 TVATNMAGRGTDIKLGKGVSELGGLAVIGTERMKSQRMQLRGRSGRQGDIGFSQFFVS 544
 T+ATNMAGRGTDIKLG+GV ELGGLAV+GTER +S+R+D QLRGRSGRQGD G +QF++S
 Sbjct: 481 TIATNMAGRGTDIKLGEGVKELGGLAVVGTERHESRRIDNQLRGRSGRQGDGPITQFYLS 540

Query: 545 FEDDLMIESGPKWAQDYFRKNRDKVNPEKPKALGQRRFQKLFQQTQEASDGKGESARSQT 604
 ED+LM G + D+ + + + + +Q+ +G +R Q
 Sbjct: 541 MEDELMRRFGAERTMAML---DRFGMDDSTPIQSKMVSRAVESSQKRVEGNNFDSRKQL 596

Query: 605 IEFDSSVQLQREYVYRERNALINGESGHFSRQIIDTVISSFI----AYLDGEVKEEL 659
 +++D ++ QRE +Y++R +I+ E + R+I++ +I S + AY E EE
 Sbjct: 597 LQYDDVLRQQREVIYKQRFVIDSE----NLREIVENMIKSSLERAIAAYTPREELPEE- 651

Query: 660 IFEVNRFI-FDNMSYNLQGISKEMSL--EIKNYLFKIADEILREKHNLLGDSFG----- 711
 ++++ + N +Y +G ++ + +E L I D I+ K+N + FG
 Sbjct: 652 -WKLDGLVDLINTTYLDEGALEKSDIFGKEPDEMLELIMDRII-TKYNEKEEQFGKEQMR 709

Query: 712 DFERTAALKAIDEAWIEVDYLQQLRTVATARQTAQRNPVFEYHKEAYKSYNIMKKEIRE 771
 +FE+ L+A+D W++ +D + QLR R AQ NP+ EY E + + M + I +
 Sbjct: 710 EFEKIVILRAVDSKMDHIDAMDQLRQGIHLRAYAQTNPRLREYQMEGFAMFEHMIESIED 769

Query: 772 QTFRNLLLSEVSFN 785
 + + ++ +E+ N
 Sbjct: 770 EVAKFVMKAEIENN 783

There is also homology to SEQ ID 3620.

SEQ ID 3978 (GBS425) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 3; MW 91kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 2; MW 116kDa).

GBS425-GST was purified as shown in Figure 220, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1295

A DNA sequence (GBSx1373) was identified in *S.galactiae* <SEQ ID 3979> which encodes the amino acid sequence <SEQ ID 3980>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3827(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

-1433-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1296

- 5 A DNA sequence (GBSx1374) was identified in *S.agalactiae* <SEQ ID 3981> which encodes the amino acid sequence <SEQ ID 3982>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2683(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 A related GBS nucleic acid sequence <SEQ ID 10001> which encodes amino acid sequence <SEQ ID 10002> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1297

A DNA sequence (GBSx1375) was identified in *S.agalactiae* <SEQ ID 3983> which encodes the amino acid sequence <SEQ ID 3984>. Analysis of this protein sequence reveals the following:

25 Possible site: 31
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.5410(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1298

A DNA sequence (GBSx1376) was identified in *S.agalactiae* <SEQ ID 3985> which encodes the amino acid sequence <SEQ ID 3986>. This protein is predicted to be preprotein translocase secY subunit. Analysis of this protein sequence reveals the following:

40 Possible site: 59
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -9.92 Transmembrane 287 - 303 (278 - 309)
INTEGRAL Likelihood = -9.08 Transmembrane 191 - 207 (186 - 210)
INTEGRAL Likelihood = -8.44 Transmembrane 104 - 120 (101 - 123)
45 INTEGRAL Likelihood = -8.23 Transmembrane 11 - 27 (9 - 41)

-1434-

5
 INTEGRAL Likelihood = -3.93 Transmembrane 133 - 149 (129 - 150)
 INTEGRAL Likelihood = -3.19 Transmembrane 347 - 363 (344 - 364)
 INTEGRAL Likelihood = -2.97 Transmembrane 158 - 174 (155 - 174)
 INTEGRAL Likelihood = -1.54 Transmembrane 246 - 262 (245 - 262)
 INTEGRAL Likelihood = -0.90 Transmembrane 372 - 388 (372 - 388)
 INTEGRAL Likelihood = -0.85 Transmembrane 64 - 80 (64 - 81)

----- Final Results -----

10
 bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15
 >GP:AAF30659 GB:AE002122 preprotein translocase [Ureaplasma urealyticum]
 Identities = 105/422 (24%), Positives = 213/422 (49%), Gaps = 49/422 (11%)

20
 Query: 2 KLLYIFEKNIILRKILITFSLIIIFLLGRYVPIPGVLISAYKQDNNFATLYSTVTGGL 61
 +LL IF+ +L +++T S++I+ +G +P+P + ++ G +F ++ + + GG L
 Sbjct: 13 QLLMIFKNNKKVLVALIVTLILFRIGSVIPMPYIKLNGNFGNQGSFFSIINLLGGGGL 72

25
 Query: 62 SQVGVFSLGIGPMMTTILLRLFT-----IGKYSSGVSQKVQFRQNVVMLVIAII 112
 SQ +F++GIGP +T I+++L + + K +K++ + ++ L +A++
 Sbjct: 73 SQFSLFAIGIGPYITAQIIMQLLSELVPPLAKLSKSGERGRKKIEVITR-IITLPLAVM 131

30
 Query: 113 QGLAITISFYHNGFSL-----TKLLLATMI--LVTGAYIISWIGNLNAEYGF- 159
 Q + I NGF + L T I +V G YI ++ +L ++ G G
 Sbjct: 132 QAVIILNLMTRANGFISIVSNAPFAIGSPLFYVYIIFLMVGTTYISLFLADLISKKG VGN 191

35
 Query: 160 GMTILVVVGMLVVGQFNNIPLIFELF-----QDGYQLAIIILFLLWTLVAMYLMITFERSE 213
 G+T+L++ G++ FN+ IF + + IL++L+ ++ + ++ S
 Sbjct: 192 GITLLILTGVASLNFHFIIFSNLGSLSKVSQIIGFILYILFYIMILIGVVFVNNST 251

40
 Query: 214 YRIPVMRTS-----IHNRLVDDAYMPIKVNASGGMAFMYVYTLIMFPQYIIILLRSIFPT 268
 +IPV +T H +L. ++PIK+ +G M ++ ++L P + L
 Sbjct: 252 RKIPVQQTGOALILDHEKL---PFLPIKIMTAGVMPVIFASSVLAIPAQVAEFLDK---Q 305

45
 Query: 269 NPDITSYNDYFSLSSIQGVVIYIMLMLVLSVAFTFVNIDPTKISEAMRESGDFIPNYRPG 328
 + ++YF + S G+ IY++L+L+ + F++V ++P K++E ++++G FIP + G
 Sbjct: 306 SMGYVVIHNYFIVDSWTGLAIYVVLILLFTFFFSYVQLNPPKMAEDIKKAGRFPVPGVQVG 365

50
 Query: 329 KETQSYLSKICYLFGTFSGFFMAFLGGVPLLFAFGNDDL- - - - -TVSSMTGIFMM 379
 +T+ +++K+ Y +AFL +P L AL + T+ T I +M
 Sbjct: 366 MDTEKHITKVIYRVNWIAPILAFACPLPHLVALVAKTINHGIPVIOPTIFGGTSIIM 425

55
 Query: 380 IT 381
 +T
 Sbjct: 426 VT 427

There is also homology to SEQ ID 3988.

50 A related GBS gene <SEQ ID 8783> and protein <SEQ ID 8784> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: 6.32
 GvH: Signal Score (-7.5): -4.07
 Possible site: 59

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 10 value: -9.92 threshold: 0.0

60
 INTEGRAL Likelihood = -9.92 Transmembrane 287 - 303 (278 - 309)
 INTEGRAL Likelihood = -9.08 Transmembrane 191 - 207 (186 - 210)
 INTEGRAL Likelihood = -8.44 Transmembrane 104 - 120 (101 - 123)
 INTEGRAL Likelihood = -8.23 Transmembrane 11 - 27 (9 - 41)
 INTEGRAL Likelihood = -3.93 Transmembrane 133 - 149 (129 - 150)
 INTEGRAL Likelihood = -3.19 Transmembrane 347 - 363 (344 - 364)
 INTEGRAL Likelihood = -2.97 Transmembrane 158 - 174 (155 - 174)

INTEGRAL	Likelihood = -1.54	Transmembrane	246 - 262 (245 - 262)
INTEGRAL	Likelihood = -0.90	Transmembrane	372 - 388 (372 - 388)
INTEGRAL	Likelihood = -0.85	Transmembrane	64 - 80 (64 - 81)
PERIPHERAL	Likelihood = 8.65	28	
5 modified ALOM score: 2.48			

----- Final Results -----

```

10      bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

```

15 ORF02350(316 - 1500 of 1827)
EGAD|6621|6420(8 - 426 of 431) preprotein translocase secY subunit {Bacillus sp.}
SP|P38375|SECY_BACHD PREPROTEIN TRANSLOCASE SECY SUBUNIT. GP|484251|dbj|EAA01191.1|D10360
secretion protein Y {Bacillus sp.} PIR|B44859|B44859 preprotein translocase secY - Bacillus
20 sp.
%Match = 12.1
%Identity = 26.8 %Similarity = 55.4
Matches = 109 Mismatches = 165 Conservative Sub.s = 116

```

```

      297       327       357       387             441       471       501
REQIDREREIPLKLLYIFEKNILRKILITFSLIIIFLLGRYPVPGV--LISAYKGQDNNFATLYSTVTGGNLSQGVGF
      || :||::|:::||::|| : : || | :| || :|
30 MFTISINPRVGDLRRKVIFTLLMLIVFRIGSFIPVPGTNRVELDFVDQANAFGL-NTFGGALGNFSIF
          10           20           30           40           50           60           70

```

35 531 582 594 624 654 681 699
SLGIGPMNTTMIILLRLF---TIGKYSSGSVSQ-----KVQQFRQNVMLVIAIIQGLAITSFQ-YHNGF---SLTKLL
::|||:| |::|: :||: :||: |||::| :|: |::|
AMGIMPYITASISVMQLLQMDVVPKFAEWAKEGEAGRRKLAQFTRYGTIVVLGFIQALGMSVGFNNFFPGLIPNSVSVYL
80 90 100 110 120 130 140 150

40

729	759	786	816	846	870	888	918
LATMILVTGAYIIISWIGNLNABEYFGF-GMTILVVVGMVLVGQFNNIPLIFEL-FQD-GYQL----AIIILFLLWTLVAMVLM							
:	:	:	:	:	:	:	:
FIALVLTAGTAFLMWLGBOITAKGVGNGISIIIFAGIAAGIPNGLNIYSTRIQDAGEFLNIVVILLLALAILAIIVG							
160	170	180	190	200	210	220	230

```

45      966      1023      1053      1083      1113      1143
ITPERSEYR-IPVM--RTSIHRLVDDA-YMPIKVNASGGMAFMVYVTLMEFPQYIIILLRSIFPTNPDITSYNDYFSL
: | : | || | | : : ::|||: | : : :||: | : | | | : ||
VIFVQQLRKIPVQYAKRLVGRNPVGGQSTHLPLKVNAAGVIPVIFALSLLIFFPTVAGLFGSDHFPVAAWVIETFDY---
      240      250      260      270      280      290      300

```

1173 1203 1233 1263 1293 1323 1353 1383
SSIQGVVIYMILMLVLSVAFTFVNIDPTKISEAMRESGDFIPNYPGKETQSYLSKICYLFGTSGGFFMAFLGGVPLLFA
: : | : : | : : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
55 THLIGMAVYALRIIGFTYFYAFIQVNPERMAENLKKQGGYIPGIRPGKATQYITPILYRLTFVGSFLFVAVLAVLPVFF-

```

1413      1440      1470      1500      1530      1560      1590      1620
LGNDLRTVSSMTGI-FMMITGMSFMILDEFOVIRIRKQYTSVFENEEN*CFILFHLGTMKIVLGMIIITCGISSRLMSV
:  ||      : |  :: : |:: : : : |:: |
IKFADLPQAIQIGGTGLLIVVGVALDTMKQIEAQLIKRSYKGFIK
      400      410      420      430

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1436-

Example 1299

A DNA sequence (GBSx1377) was identified in *S.agalactiae* <SEQ ID 3989> which encodes the amino acid sequence <SEQ ID 3990>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3002(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF61315 GB:U96166 unknown [Streptococcus cristatus]
 Identities = 30/78 (38%), Positives = 41/78 (52%)

Query: 276 ALTVTLTDDIWELEHLLQRCPTDFHIAAPVYCSDRKQLVGYPNYLHEAITEEQFEVL 335
 AL +T +D + ++E LL + PN FHI A S L L YPN L+ I + L
 Sbjct: 289 ALILTNSDQLEQIEQLLTQLPNVHFGAITEMSGHLMGLNRYPNVSLYPNIRPAKVAEL 348

Query: 336 LLNSDIYLDINHGEVWN 353
 D+YLDIN +E+ N
 Sbjct: 349 FERCDLYLDINISDEILN 366

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1300

A DNA sequence (GBSx1378) was identified in *S.agalactiae* <SEQ ID 3991> which encodes the amino acid sequence <SEQ ID 3992>. This protein is predicted to be eps7. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> May be a lipoprotein

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAC07458 GB:AX009404 product = eps7 [Streptococcus thermophilus]
 Identities = 87/232 (37%), Positives = 133/232 (56%), Gaps = 22/232 (9%)

Query: 10 VSVIIPVYNAAPYLEGCVNTILGQTYQVFEILLIDDGSTDTSASICDQLSLRDNRI RVF 69
 +S++IPVYN Y++ C+++IL QT+ EI+L+DDGSTD S ICD S D RI+V H
 Sbjct: 3 ISIVIPVYNVDYIKKCLDSILSQTFSLEILLVDDGSTDLSGRICDYSENDKRIKVIH 62

Query: 70 IEENGASKARNFGLARISPESQFVTTFVDSDDWVKENYLEVLLAQKEKYNADIVISNYIY 129
 NGG S+ARN G+ + S+++TF+DSDD+V +Y+E L + +NADI I+++
 Sbjct: 63 TANGGQSEARNVGIKNAT--SEWITFIDSDDYVSSDYIEYLYNLIQVHNADISTASF--- 117

Query: 130 RETEDIFGYIYTDKDFV-----IEEISAQTAIDRQVHWHLNSVFIWIKLYRRELF 183
 YIT K + + + A+TAI R + LN + +WKG+YR E F+
 Sbjct: 118 -----TYITPKKIIKHGNGEVALMDAKTAIRRL---LNEGFDMGVWGMRYRTEYFN 166

Query: 184 TITFPIDKVFEDLVSVLLFIKSKKTILVNGSYGYRIRPNSIMTSAFSSKR 235
 F K+FED L++ +F ++ + Y Y R NS + F+ K+

-1437-

Sbjct: 167 KYKFVSGKLFEDSLITYQIFSEASTIVFGAKDIYFVNRKNSTVNGTFNIKK 218

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 1301

A DNA sequence (GBSx1379) was identified in *S.agalactiae* <SEQ ID 3993> which encodes the amino acid sequence <SEQ ID 3994>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.1569(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 1302

A DNA sequence (GBSx1380) was identified in *S.agalactiae* <SEQ ID 3995> which encodes the amino acid sequence <SEQ ID 3996>. Analysis of this protein sequence reveals the following:

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.1662(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 Example 1303

A DNA sequence (GBSx1381) was identified in *S.agalactiae* <SEQ ID 3997> which encodes the amino acid sequence <SEQ ID 3998>. This protein is predicted to be a glycosyl transferase (gspA). Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.2606(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

-1438-

>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus ducreyi]
Identities = 62/177 (35%), Positives = 105/177 (59%), Gaps = 8/177 (4%)

5 Query: 3 YARYIPQLIDAEKVLVLDIDTLVVDNLDKLFIEIELGDYPIAAILD--GDGIY-----FN 55
+ RY+I I+ +KV+YLD D +V +L +L++ ++ +Y +AA+ D + IY FN
Sbjct: 89 FFRYFISDFIEQDKVIYLDADIVVNGSLTELYQTDISNYFLAAVKDIISEKIYVNNHIFN 148

10 Query: 56 SGVMLINSLYWMRYRVTEKLLITERELDNGIFGDQGVNLNLLFDNNWLKLEDKYNAQVGN 115
+G++LIN+ W + +T+ L ++E+ +++ DQ +LNL+F + WLKL YN +G
Sbjct: 149 AGMLLINNKKWREHNITQFCLSLSEKYINSLPDADQSILNLFKDKWLKLNRGYNYLIGT 208

Query: 116 DLGAFYENWQGYFDRNFES-PTIIHYCTHDKPWNITFSSSRFRETWQYEQLDWNEVF 171
D F Y + E+ P IIHY T KPW ++RFR +W Y +L+W +++
15 Sbjct: 209 DYLFKYGKTRYLEDLGETIPLIIHYNTEAKPWLNIENFRFRNIYWFYELNWQDIY 265

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 1304

A DNA sequence (GBSx1384) was identified in *S.agalactiae* <SEQ ID 3999> which encodes the amino acid sequence <SEQ ID 4000>. This protein is predicted to be a glycosyl transferase. Analysis of this protein sequence reveals the following:

25 Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1157(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus ducreyi]
35 Identities = 103/259 (39%), Positives = 156/259 (59%), Gaps = 3/259 (1%)

Query: 7 IALAADFGYQEYKTIKISICPHNQFIDFYILNDDFPVVEWFQMMYHLSKMDCTISNTKI 66
I LAA+ Y E + T IKSI HN+ I FY+LN D+P EWF ++ L K++ I + K+
40 Sbjct: 10 IVLAANQSYSEYILTTIKSIYLNHKNHIFRYLLNRDYPTEWFDILNKLRLNSELIDIKV 69

Query: 67 FNEEIKHFK-FQKPMFYPTYFRYFIPEVIHEDKVLVLDCEMIITSDLTISFTLDISKYGV 125
N+ IK+FK + T+FRYFI + I +DKV+YLD D+++ LT ++ DIS Y +
Sbjct: 70 TNDTIKNFKTYSHISSDTTFFRYFISDFIEQDKVIYLDADIVVNGSLTELYQTDISNYFL 129

45 Query: 126 AAVRDDLLEEDGKEDYFNSGLLLINNIWFREQGISQRLDQYHNDQDVLND 185
AAV+D + E+ FN+G+LLINN WRE I+Q L + + +L DQ +LN
Sbjct: 130 AAVKDIISEKIYVNNHIFNAGMLLINNKKWREHNITQFCLSLSEKYINSLPDADQSILNL 189

50 Query: 186 VLCDNWLELDETYNYHTGADMLYNLFQQSERQLNRRKDLPKVIHY-TATKPKYLETSVR 244
+ D WL+L+ YNY G D L+ + ++ + + +P +IHY T KPW + + R
Sbjct: 190 IPKDKWLKLNRGYNYLIGTDYLFKYGKTRYLEDLGETIPLIIHYNTEAKPWLNI-FNTR 248

Query: 245 WRDIWWEYNRLEWRDIFTR 263
+R+I+W Y L W+DI+ +
55 Sbjct: 249 FRNIYWFYELNWQDIYAK 267

No corresponding DNA sequence was identified in *S.pyogenes*.

-1439-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1305

5 A DNA sequence (GBSx1385) was identified in *S.agalactiae* <SEQ ID 4001> which encodes the amino acid sequence <SEQ ID 4002>. This protein is predicted to be a glycosyl transferase. Analysis of this protein sequence reveals the following:

Possible site: 28
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2679(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus
ducreyi]
Identities = 94/263 (35%), Positives = 158/263 (59%), Gaps = 4/263 (1%)

20 Query: 2 KKTIVLGADFQYRDQVMTTIKSIVSHNQHLTIYIINTDFPVEWFNINHSLEQFDCRVKN 61
K IVL A+ Y + ++TTIKSI HN+H+ Y++N D+P EWF+ILN+ L + + + +
Sbjct: 7 KMNIVLAANQSYSEYILTITIKSIYLNKHIRFYLLNRDYPTEWFDILNNKLRKLNSEIID 66

25 Query: 62 IPISSDVFEIGIPTLSHISV-AGFFRFWFIPIHLEEEIVLYLDSIVIRGSLDPLFDINLEE 120
I +++D + T SHIS FFR+FI +E++ V+YLD+D++V GSL L+ ++
Sbjct: 67 IKVTNDTIKNFKTYSHISSDTFFRYFISDFIEQDKVIYLDADIVNGSLTELYQTDISN 126

30 Query: 121 NLLGAVADHFSTLYYGDTPVFSFNSGVMLINNSLWKKEEIVNSLMRIADKG-SAVGVGDQ 179
L AV D S Y + FN+G++LINN W++ I + +++K +++ DQ
Sbjct: 127 YFLAAVKDIISEKIYVNNH--IFNAGMLLINNKKWREHNTQFCLSLSEKYINSLPDADQ 184

35 Query: 180 EYLNILTQNRWIDIGQYNVQIGQDVNINAYGRPDLYHFYDDCEPVIVHYSQDKPWNKY 239
LN++ +++W+ + + YN IG D YG+ + P+I+HYN++ KPW
Sbjct: 185 SILNLIKFDKWLKLNRGVNYLIGTDYLFYKYGKTRYLEDLGETIPLIIHYNTEAKPWLNI 244

Query: 240 SQSRYRSEWYWFYFGLWSVYIAQ 262
+R+R+ +W+Y+ L W IYA+
Sbjct: 245 FNTIRFRNIYWFYFELNWQDIYAK 267

40 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1306

45 A DNA sequence (GBSx1386) was identified in *S.agalactiae* <SEQ ID 4003> which encodes the amino acid sequence <SEQ ID 4004>. Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2996(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1440-

A related GBS nucleic acid sequence <SEQ ID 10003> which encodes amino acid sequence <SEQ ID 10004> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAC75095 GB:AE000294 putative Galf transferase [Escherichia coli K12]
    Identities = 68/286 (23%), Positives = 122/286 (41%), Gaps = 18/286 (6%)

    Query: 77 STRMDGIIAGLGRGDIVVFQVPTWNSTEFDELFLDKLQAYGARIITFVHDLVPLMFESNF 136
              S ++ + GL D+++F P F +L + RI+ +HDI L
    10  Sbjct: 50 SVKLSTFLCGLNKDVLIFNFPMAKPFWHILSFFHRLK--RIVPLIHDIDELRGGGGS 107

    Query: 137 YLLDRVIDMYNRSDVILPTKAMHDYLIKGMTTSKVLVQEVDHPVNIIDLPREPEC---Q 193
              D V D+VI M YL K M+ K+ +++D+ V+ D+ + Q
    15  Sbjct: 108 ---DSV--RLATCDMVISHNPQMTKYL-SKYMSQDKIKDIFDYLVSSDEVHRDVTDKQ 161

    Query: 194 KVLFSFAGDIQRFPFVNDWKENIPLIYYGDSRLNSEANVHAQGWKDDVELMLSLSKRG-G 252
              + + +AG++ R + E +G ++ N G D + ++ G
    20  Sbjct: 162 RGVIIYAGNLSRHKCSFIYTEGCDFTLFG--VNYENKDNPKYLG-SFDAQSPEKINLPGMQ 218

    Query: 253 FGLCWSEDREELVERR---YSRMNASYKLSSTFLAAGLPITIANHDISSRDFIKQHGLGFTV 309
              FGL W D E Y + N +K S +L+ LP+ + DFI + +G+ V
    25  Sbjct: 219 FGLIWDGDSVETCSGAFGDYLFNPNPKTSLYLSMELPVFIWDKAALADFIVDNRIGYAV 278

    Query: 310 ETLEEAVEKINMEKETYSYVENVEKIATLLRNGYITKLLIDAV 355
              +++E E +++M ETY EN + I+ +R G + +L + +
    25  Sbjct: 279 GSIKEMQEIVDSMTIETKQISENTKIISQKIRTGSYFRDVL EEVI 324
  
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 1307

A DNA sequence (GBSx1387) was identified in *S.agalactiae* <SEQ ID 4005> which encodes the amino acid sequence <SEQ ID 4006>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 33
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3098(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    40  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  
```

The protein has homology with the following sequences in the GENPEPT database.

```

45  >GP:AAA73093 GB:M76233 [Rabbit smooth muscle myosin light chain
    kinase mRNA, complete CDS.], gene product [Oryctolagus
    cuniculus]
    Identities = 23/63 (36%), Positives = 36/63 (56%)

    Query: 5 QPAPALQVRVQCQPAPVLQPVPRCPALALQVRVQCQPAQVLQQVPRCPAQVLQQVPRC 64
              +PA L+ V +PA L+PV +PA L+ V +PA+ L+ V +PA+ L+ V
    50  Sbjct: 225 KPAETLKPVGNAKPAETLKPVGNAKPAETLKPVGNAKPAETLKPVGNAKPAETLKAVANA 284

    Query: 65 QPA 67
              +PA
    Sbjct: 285 KPA 287
  
```

55 No corresponding DNA sequence was identified in *S.pyogenes*.

-1441-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1308

A DNA sequence (GBSx1388) was identified in *S.agalactiae* <SEQ ID 4007> which encodes the amino acid sequence <SEQ ID 4008>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -9.24    Transmembrane  189 - 205 ( 173 - 245)
    INTEGRAL    Likelihood = -9.24    Transmembrane  213 - 229 ( 206 - 245)
10    INTEGRAL    Likelihood = -7.96    Transmembrane   95 - 111 (  83 - 185)
    INTEGRAL    Likelihood = -7.96    Transmembrane  115 - 131 ( 112 - 185)
    INTEGRAL    Likelihood = -7.96    Transmembrane  135 - 151 ( 132 - 185)
    INTEGRAL    Likelihood = -7.96    Transmembrane  155 - 171 ( 152 - 185)
    INTEGRAL    Likelihood = -6.85    Transmembrane   15 -  31 (   8 -  45)
15    INTEGRAL    Likelihood = -4.09    Transmembrane   39 -  55 (  35 -  57)
    INTEGRAL    Likelihood = -4.09    Transmembrane   63 -  79 (  59 -  81)
    INTEGRAL    Likelihood = -2.71    Transmembrane  235 - 251 ( 235 - 251)
    INTEGRAL    Likelihood = -0.11    Transmembrane  253 - 269 ( 253 - 269)

20    ----- Final Results -----
        bacterial membrane --- Certainty=0.4694(Affirmative) < succ>
        bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

25 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC16164 GB:AF010496 ice nucleation protein [Rhodobacter capsulatus]
  Identities = 85/286 (29%), Positives = 119/286 (40%), Gaps = 17/286 (5%)

Query: 3  ALVLADVDALVETLVLADVVALIEALVLADIEALV----EALVLADIEALVEALVLADID 58
30      AL  A  AL  T +    A ++ L AD+  L    +AL  A I AL  + + A
Sbjct: 523 ALSDAQAGALTSTQIGLLSTAANKGLSTADMAGLTTAEQAALTSQAIAALSSSQIRAMTT 582

Query: 59  ALVEALVLADIEALVEALVL----ADIDALVEALVLADVEALIEALVLALVEALVLADVE 114
35      A + AL  A I+ L  + +L    ADI AL    A  + I AL  +LV A+  AD+
Sbjct: 583 AQIAALGTAQIKGLTASNILGLETADIVALTTTQAPALSSSQIAALSTSLVAAMETADLA 642

Query: 115 ALIEALVLAL----VEALVLADVEAL----IEALVLALVEALVLADVEALIEALVLALVE 166
40      L  A          + AL  A  A+    I  + A ++ L AD+ AL  A  + +
Sbjct: 643 KLSAATFKGFSSTQITALTTAQAGAIGTDQIAQITTAIKGLESADIAALANATLAKMTT 702

Query: 167 ALVLADVEALIEALVLADVD-ALVLALVEALVLALVEALILAEVEALVLALVEALVLALV 225
45      A V    A + L  ++ L  A V+AL  A + L  ++ AL    AL    V
Sbjct: 703 AQVAVLGSAQLTGLTTTQINTVLTTAQVKALGAAALAGLGTDDIVALTGQAAALSSSTQV 762

Query: 226 EALILALVEALVLADVDALMEALVLADVEALMEALVLADVDALVEA 271
50      AL  A + AL  AD  AL  A +    + AL  +DAL  A
Sbjct: 763 AALSTAQISALQTADFAALSTAAIKGLSSTQITALSTGQIDALTTA 808

```

No corresponding DNA sequence was identified in *S.pyogenes*.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1309

A DNA sequence (GBSx1389) was identified in *S.agalactiae* <SEQ ID 4009> which encodes the amino acid sequence <SEQ ID 4010>. Analysis of this protein sequence reveals the following:

```

55    Possible site: 41
    >>> Seems to have no N-terminal signal sequence

```

-1442-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2297(Affirmative) < succ>
 bacterial membranes --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 10 vaccines or diagnostics.

Example 1310

A DNA sequence (GBSx1390) was identified in *S.galactiae* <SEQ ID 4011> which encodes the amino acid sequence <SEQ ID 4012>. This protein is predicted to be fimbriae-associated protein Fap1. Analysis of this protein sequence reveals the following:

15 Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.3138(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:BAA97453 GB:AB029393 streptococcal hemagglutinin [Streptococcus gordonii]
 Identities = 388/968 (40%), Positives = 518/968 (53%), Gaps = 68/968 (7%)

30 Query: 13 VDTKSRVKMHKSEKNWVRTVMSHFNLKFAIKGRATVEADVCIQDVEKEDRLSSGNLTYLK 72
 V+ +R K+ KS K+W+R S F L + +KG +V V +E + G L YLK
 Sbjct: 13 VERVTRFKLIKSGKHWLRAATSQFGLLRMLMGADISSVEV---KVAEEQSVKEGGLNYLK 69

Query: 73 GILAAGALVGGASLTSR-VYADETPVVQEQSSSVPTLAEQTEVTV--KTTTVQNHQDGTV 129
 GI+A GA++GGA +TS VYA+E +++ + LA + E + + T + +
 Sbjct: 70 GIIATGAVLGGAVVTSSSVYAEQQALEKVIDTRDVLATRGEAVLSEEAATLSSEGANP 129

35 Query: 130 SKNIIDNSVSMSESASTSTSESVMMSGSTLTLSVSESVSTLSALTSASESISTASASESV 189
 +++ D+ S S S SA+ S S S+S+S S S S S S S+S S+SES S S S SV
 Sbjct: 130 VESLSDTLASASASAN-SVSTSISSISSEFSVSASASLSSSSSLQSSSESASASELSV 188

40 Query: 190 SKSTSISEVSNILETQASLTDKGRESFSANQIVTESSLVTDAGKNASVSSLEITKPKSE 249
 S STS S S TQ+S + S S+N + T S V+ +NA V + + +E
 Sbjct: 189 SASTSQSFSSSTTSSTQSSNNESLISSDSSNSLNTNQS-VSARNQNARVTRRAVAANDTE 247

45 Query: 250 LQTSKMSNESLITPEKSQVMIASDKTGNESLTPTIRLKSVIQPRSMNMLTSSSEMDLIPL 309
 K + + E + ++ T N + ++ N+ ++ L P
 Sbjct: 248 APQVKSGDYVVYRGESFEYY--AEITDNSGQVNRVVIR-----NVEGGANSTYLSNP 297

Query: 310 EEVSDTEMLGKDVSELQKVNIALKDNTLSEPGTVKLDSSSENVLNLFASVNEGDFV 369
 TE LG+ ++ +Q L+ E ++ + ++ + +A G+
 Sbjct: 298 WVKYSTENLGRPGNATVQN---PLRTRIFGEVPLNEIVNEKSYTRYI--VAWDPSGN-- 350

50 Query: 370 TVKLSNLDTOGIGITILKVQDIMDETQQLLATGSYSPLTHNITY-----TWTRYAST 421
 ++ DN + G+ + +E Y P ++TY T R A
 Sbjct: 351 ATQMDVNDANRGLERFVLTVKSQNE-----KYDPAESSVTYVNNLSNLSTSEREAVA 402

55 Query: 422 LNNIKARVNMPVWPDQRI-----ISKTTSDKQCFTATLNNQVASIE---ERVQYNPS 471
 A N+P P +I ++ T DK T N V ++ S S
 Sbjct: 403 AAVRAANPNIP--PTAKITVSQNGTVTITYPDKSTDITIPANRVVKDLQISKNSASQSSS 460

Query: 472 VTEHTNVKTNVRSRIMKLDDEKQTEYITQINPECKEMYFASGLGNLYTIIGSDGTSISP 531

-1443-

V+ + T+V + I ++ + + ++ S+ S S
 Sbjct: 461 VSASQSASTSVSASI---SASMSASVSVSTASASTSASVSASESASTSASVSASESASTS- 516
 Query: 532 VNLLNAEVKILKINSKNLIDSMQNDSPFEDVTSQYSYTNDSKITIDWKTNISISTT 591
 5 A V K++S + + S ++ + + S + S + S+S++T
 Sbjct: 517 -----ASVGASKSSSTASVSASESASTSASVSASESASTSASVSASESASTSASVSAST 571
 Query: 592 SYVVLVKIPKQSGVLYSTVSDINQTYGSKYSYGHNTNISGDSANAEIKL-LSESASTSAS 650
 10 S + ST + ++ + + S ++S A+ + SESASTSAS
 Sbjct: 572 SASTSASVSASESA--STSASVSASESASTS---ASVSASESASTSASVSASESASTSAS 626
 Query: 651 TSASTSASMSASTSASTSASMSASTSASTSASTSASMSASTSASTSASTSASTSASTSAS 710
 SAS S+S SAS SAS SAS SAS SAS SASTSAS+SASTSASTSAS SASTSASTSAS
 15 Sbjct: 627 VSASESSSTASVSASESASTSASVSASESASTSASVSASTSASTSASVSASTSASTSAS 686
 Query: 711 MSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 770
 +SASTSASTSAS SAS SASTSAS SAS SASTSAS SASTSASTSAS+SASTSASTSAS
 Sbjct: 687 VSASTSASTSASVSASESASTSASVSASESASTSASVSASTSASTSASTSASTSASTSAS 746
 Query: 771 TSASTSASMSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 830
 SAS SAS SAS SASTSASTSAS SAS SASTSAS SAST ASTSAS+SAS SASTSAS
 20 Sbjct: 747 VSASESASTSASVSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 806
 Query: 831 TSASMSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 890
 25 SAS SASTSAS SAS SASTSAS SAS SASTSAS SAS SASTSAS SAS SASTSAS
 Sbjct: 807 VSASESASTSASVSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS 866
 Query: 891 MSATTSASTSVSTASASTSASTSASTSSSSSVTNSSSKEKVYALPSTGDDQDYSVTATAG 950
 +SA+TSASTS S SAS SASTSAS S+S S ++++S SA S +T+
 30 Sbjct: 867 VSASTSASTSASVSASESASTSASVSASESASTSASTSASTSASTSASTSASTSASTSAS 926
 Query: 951 LGLMTGAT 958
 + T A+
 35 Sbjct: 927 VSASTSAS 934

There is also homology to SEQ ID 760.

SEQ ID 4012 (GBS68) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 4; MW 131.2kDa).

40 GBS68d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 153 (lane 14; MW 103kDa) and in Figure 239 (lane 13; MW 103kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 17; MW 78kDa), in Figure 153 (lane 17; MW >78kDa) and in Figure 184 (lane 10; MW 78kDa). Purified GBS68d-GST is shown in Figure 246, lane 5.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1311

A DNA sequence (GBSx1391) was identified in *S.agalactiae* <SEQ ID 4013> which encodes the amino acid sequence <SEQ ID 4014>. This protein is predicted to be RofA. Analysis of this protein sequence reveals the following:

50 Possible site: 30
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 55 bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1444-

A related GBS nucleic acid sequence <SEQ ID 10005> which encodes amino acid sequence <SEQ ID 10006> was also identified.

There is also homology to SEQ ID 3750.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1312

- A DNA sequence (GBSx1392) was identified in *S.agalactiae* <SEQ ID 4015> which encodes the amino acid sequence <SEQ ID 4016>. This protein is predicted to be Nra. Analysis of this protein sequence
10 reveals the following:

Possible site: 16
>>> Seems to have a cleavable N-term signal seq.

- 15 ----- Final Results -----
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 3750.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1313

A DNA sequence (GBSx1393) was identified in *S.agalactiae* <SEQ ID 4017> which encodes the amino acid sequence <SEQ ID 4018>. Analysis of this protein sequence reveals the following:

- 25 Possible site: 19
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
30 bacterial cytoplasm --- Certainty=0.3674(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 35 >GP:AAA27020 GB:M80215 uvs402 protein [Streptococcus pneumoniae]
Identities = 577/663 (87%), Positives = 633/663 (95%), Gaps = 1/663 (0%)
Query: 1 MIDRKDTNRFKLVSKYSPSGDQPQAIEITLVDNIEGGEKAQILKGATGTGKTYTMSQVIAQ 60
MI+ N+FKLVSKY PSGDQPQAIE LVDNIEGGEKAQIL GATGTGKTYTMSQVI++
Sbjct: 7 MINHITDNQFKLVSKYQPSGDQPQAIEQLVDNIEGGEKAQILMGATGTGKTYTMSQVISK 66
40 Query: 61 VNKPTLVIAHNKTLAQQLYGEFKEFFPDNAVEYFVSYYDYYQPEAYVPSSDTYIEKDSSV 120
VNKPTLVIAHNKTLAQQLYGEFKEFFP+NAVEYFVSYYDYYQPEAYVPSSDTYIEKDSSV
Sbjct: 67 VNKPTLVIAHNKTLAQQLYGEFKEFFPENAVEYFVSYYDYYQPEAYVPSSDTYIEKDSSV 126
45 Query: 121 NDEIDKLRSATSSLLERNDVIVVASVSCIYGLGSPKEYADSVVSLRPGQEISRDQLLN 180
NDEIDKLRSATS+LLERNDVIVVASVSCIYGLGSPKEYADSVVSLRPG EISRD+LLN+
Sbjct: 127 NDEIDKLRSATSALLERNDVIVVASVSCIYGLGSPKEYADSVVSLRPGLEISRDKLLND 186
50 Query: 181 LVDIQFERNDIDFQGRGFRVRGVDVVEFPASRDEHAFRIEFFGDEIDRIREIESLTGRVL 240
LVDIQFERNDIDFQRG+FRVRGVDVVE+FPASRDEHAFR+EFFGDEIDRIRE+E+LTG+VL
Sbjct: 187 LVDIQFERNDIDFQGRGFRVRGVDVVEIFPASRDEHAFRVEFFGDEIDRIREVEALTGQVL 246

-1445-

Query: 241 GEVEHLAIFPATHFMTNDEHMEEAISKIQAEMENQVELFEKEGKLEAQRIRQRTEYDIE 300
 GEV+HLAIFPATHF+TND+HME AI+KIQAE+E Q+ +FEKEGKL+EAQR++QRTEYDIE
 Sbjct: 247 GEVDHLAIFPATHFVTNDDHMEVAIAKIQAELEEQLAVFEKEGKLEAQRIRQRTEYDIE 306

5 Query: 301 MLREMGYTNGVENYSRHMDGRSEGEPPFTLLDFFPEDFLIMIDESHMTMGQIKGMYNDR 360
 MLREMGYTNGVENYSRHMDGRSEGEPP+TLLDFFP+DFLIMIDESHMTMGQIKGMYNDR
 Sbjct: 307 MLREMGYTNGVENYSRHMDGRSEGEPPYTTLLDFFPDFFLIMIDESHMTMGQIKGMYNDR 366

10 Query: 361 SRKEMLVNNGFRLPSALDNRLRREEFESHVHQIVVVSATPGDYEMEQTDTTVVEQIIRPT 420
 SRK+MLVNNGFRLPSALDNRLRREEFESHVHQIVVVSATPGDYE EQT+TV+EQIIRPT
 Sbjct: 367 SRKKMLVNNGFRLPSALDNRLRREEFESHVHQIVVVSATPGDYENEQTETVIEQIIRPT 426

15 Query: 421 GLLDPEVEVRPSMGQMDLLGEINLRTEKGERFITTTLTRMAEDLTDYKEMGVKVKYM 480
 GLLDPEVEVRP+MGQ+DDLLGEIN R EK ERTFITTLTK+MAEDLTDY KEMG+KVKYM
 Sbjct: 427 GLLDPEVEVRPTMGQIDLLGEINARVEKNERTFITTLTKMAEDLTDYFKEMGIKVKYM 486

20 Query: 481 HSDIKTLERTEIIRDLRLGVFDVLGINLLREGIDVPEVSLVAILDADKEGFLRNERGLI 540
 HSDIKTLERTEIIRDLRLGVFDVL+GINLLREGIDVPEVSLVAILDADKEGFLRNERGLI
 Sbjct: 487 HSDIKTLERTEIIRDLRLGVFDVLGINLLREGIDVPEVSLVAILDADKEGFLRNERGLI 546

25 Query: 541 QTIGRAARNSNGHVIMYADKITDSMQRAMDETARRRRLQMDYNEKHGIVPQTIKKEIRD 600
 QTIGRAARNS GHVIMYAD +T SMQRA+DETARRR++QM YNE+HGIVPQTIKKEIRD
 Sbjct: 547 QTIGRAARNSGHVIMYADTVTQSMQRAIDETARRRKIQMAYNEEHGIVPQTIKKEIRD 606

30 Query: 601 IAITKSNDSKPEKVVDSYSSLSKKERQAEIKALQQMQEAAELDFELAAQIRDVILELK 660
 IA+TK+ ++ +K VD +SL+K+ER+ +K L++QMGEA E+LDFELAAQIRD++LE+K
 Sbjct: 607 IAVTKAVAKEE-DKEVDINSINKQERKELVKKLEKQMQEAVEVLDFFELAAQIRDMMLEVK 665

Query: 661 AID 663
 A+D
 Sbjct: 666 ALD 668

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4019> which encodes the amino acid sequence <SEQ ID 4020>. Analysis of this protein sequence reveals the following:

35 Possible site: 55
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4386(Affirmative) < succ>
 40 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 570/663 (85%), Positives = 625/663 (93%)

45 Query: 1 MIDRKDTNRFKLVSKYSPSGDQPQAIE TLVDNIEGGEKAQILKGATGTGKTYTMSQVIAQ 60
 MID++D FKL SKY PSGDQPQAIE+LVDNIEGGEKAQIL GATGTGKTYTMSQVI++
 Sbjct: 1 MIDKRDDKPFKLKSKYKPSGDQPQAIESLVDNIEGGEKAQILGATGTGKTYTMSQVISK 60

50 Query: 61 VNKPTLVIAHNKTLAGQLYGEFKEFFPDNAVEYFVSYDYDYQPEAYVPSSDTYIEKDSSV 120
 VNKPTLVIAHNKTLAGQLYGEFKEFFPDNAVEYFVSYDYDYQPEAYVPSSDTYIEKDSSV
 Sbjct: 61 VNKPTLVIAHNKTLAGQLYGEFKEFFPDNAVEYFVSYDYDYQPEAYVPSSDTYIEKDSSV 120

55 Query: 121 NDEIDKLRSATSSLLERNDIVVASVSCIYGLGSPKEYADSVSLRPGQEISRDQLLN 180
 NDEIDKLRSATSSLLERNDIVVASVSCIYGLGSPKEYADS VSLRPGQEISRD LLN
 Sbjct: 121 NDEIDKLRSATSSLLERNDIVVASVSCIYGLGSPKEYADSAVSLRPGQEISRD TLLNQ 180

60 Query: 181 LVDIQFERNDIDFQRGKFRVRGVDVVEFPASRDEHAFFRFEFFGDEIDRIEIESLTGRVL 240
 LVDIQFERNDIDFQRG FRVRGVDVVEFPASRDEHAFFR+EFFGDEIDRI EIESLTG+ +
 Sbjct: 181 LVDIQFERNDIDFQRGCFRVRGVDVVEFPASRDEHAFFRVEFFGDEIDRICIESLTGKTI 240

65 Query: 241 GEVEHLAIFPATHFMTNDEHMEEAISKIQAEMENQVELFEKEGKLEAQRIRQRTEYDIE 300
 GEV+HL +FPATHF+TNDHME++I+KIQAE+ Q++LFE EGKL+EAQR+RQRTEYDIE
 Sbjct: 241 GEVDHLVLFPATHFVTNDEHMQSIAKIQAEALAEQLQFSEGKLEAQRIRQRTEYDIE 300

-1446-

Query: 301 MLREMGYTNGVENYSRHMDGRSEGEPPFTLLDFFPEDFLIMIDESHMTMGQIKGMYNGDR 360
MLREMGYT+GVENYSRHMDGRS GEPP+TLLDFFPEDFLIMIDESHMTMGQIKGMYNGD+

Sbjct: 301 MLREMGYTSVENYSRHMDGRSPGEPPYTLDDFFPEDFLIMIDESHMTMGQIKGMYNGDQ 360

5 Query: 361 SRKEMLVNYGFRPLPSALDNRLRREEFESHVHQIVYVSATPGDYEMEQTDTVVEQIIRPT 420
+RK+MLV+YGFRLPSALDNRLRREEFESHVHQIVYVSATPG+YEM QT+T++EQIIRPT

Sbjct: 361 ARKQMLVDYGFRLPSALDNRLRREEFESHVHQIVYVSATPGYEMSQTNTIIEQIIRPT 420

10 Query: 421 GLLDPEVEVRPSMQQMDLLGEINLRTEKERTFITTLTKRMAEDLTDYLKEMGVKVKYM 480
GLLDPE++VR SMQMDLLGEIN R + ERTFITTLTK+MAEDLTDYLKEMGVKVKYM

Sbjct: 421 GLLDPEIDVRSSMQQMDLLGEINQVRDERTFITTLTKRMAEDLTDYLKEMGVKVKYM 480

15 Query: 481 HSDIKTLERTEIIRDLRLGVFDVLIGINLLREGIDVPEVSLVAILDADKEGFLRNERGLI 540
HSDIKTLERTEIIRDLRLGVFDVLIGINLLREGIDVPEVSLVAILDADKEGFLRNERGLI

Sbjct: 481 HSDIKTLERTEIIRDLRLGVFDVLIGINLLREGIDVPEVSLVAILDADKEGFLRNERGLI 540

20 Query: 541 QTIGRAARNSNGHVIMYADKITDSMQRAMDETARRRRLQMDYNEKHGIVPQTIKKEIRDL 600
QTIGRAARN +GHVIMYADK+TDSMORA+DETARRR +Q+ YN+ HCIVPQTIKK+IR L

Sbjct: 541 QTIGRAARNVDGHVIMYADKMTDSMQRADDETARRREIQIAYNKAHGIVPQTIKKDIRGL 600

Query: 601 IAITKSNDSKPEKVVYSSLSKKERQAEIKALQQQMGEAAELLDFFELAAQIRDVILELK 660
I+I+K++ +D ++ +DY S+S+ ER+ I ALQ+QMGEAAELLDFFELAAQ+RD+ILELK

Sbjct: 601 ISISKTSNDISKEEMDYESMSRGERKEAINALQKQMGEAAELLDFFELAAQMRDLILELK 660

25 Query: 661 AID 663
+D

Sbjct: 661 LMD 663

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1314

A DNA sequence (GBSx1394) was identified in *Sagalactiae* <SEQ ID 4021> which encodes the amino acid sequence <SEQ ID 4022>. Analysis of this protein sequence reveals the following:

Possible site: 31

35 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.78	Transmembrane	284 - 300 (274 - 303)
INTEGRAL	Likelihood = -10.08	Transmembrane	20 - 36 (16 - 53)
INTEGRAL	Likelihood = -5.52	Transmembrane	117 - 133 (114 - 137)
INTEGRAL	Likelihood = -5.15	Transmembrane	203 - 219 (201 - 225)
40 INTEGRAL	Likelihood = -3.29	Transmembrane	183 - 199 (182 - 200)
INTEGRAL	Likelihood = -1.54	Transmembrane	74 - 90 (73 - 90)
INTEGRAL	Likelihood = -0.48	Transmembrane	37 - 53 (37 - 53)

----- Final Results -----

45 bacterial membrane --- Certainty=0.5713(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:CAA22372 GB:AL034446 putative transmembrane protein
[Streptomyces coelicolor A3(2)]
Identities = 58/190 (30%), Positives = 96/190 (50%), Gaps = 11/190 (5%)

Query: 114 GWS--IGFILFSISVITAYILGGLDFHSYDYSK-ATIFYVVTLPLFWLIQSSTEELLTRG 170
GW IGF LF +VIT G Y+V ++ + L+ F + TEE++ RG

55 Sbjct: 98 GWGTLIGFGLFG-AVITNLFASGY---YEVDGLGSVQGAIGLVGFMAAAATEEVVFRG 152

Query: 171 WLLPLINHRFHLAVAIGVSSTLFGILHLVNAHVTFSLIVSI-ICSGVLSLYMIKSGNIW 229
L +I +A+G++ +FG++HL+N T ++I I +G +++ + N+W

60 Sbjct: 153 VLFRIIEEHIGTYLALGLTGLVFGMLHLNEDATLWGALAIATEAGFMLAAAYAATRNW 212

Query: 230 SVAALHGAWNFSQGNLYGIAVSGQKAGASLLHFTVKENAPDWISGGAFGIEGSLISIFVL 289

-1447-

+H WNF+ G ++ VSG LL T+ + P ++GG FG EGS+ S+
 Sbjct: 213 LTIGVHFGWNFAAGGVFSTVVSNGDSEGLLDATM--SGPKLLTGDFGPEGSVYSGVFG 270
 Query: 290 LAATIIYLLWL 299
 + + LWL
 Sbjct: 271 VLLTLVFLWL 280

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1315

A DNA sequence (GBSx1395) was identified in *S.agalactiae* <SEQ ID 4023> which encodes the amino acid sequence <SEQ ID 4024>. This protein is predicted to be glutamine-binding periplasmic protein/glutamine transport system perme. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -8.97 Transmembrane 532 - 548 (523 - 553)
 INTEGRAL Likelihood = -7.38 Transmembrane 700 - 716 (696 - 720)
 INTEGRAL Likelihood = -4.57 Transmembrane 562 - 578 (558 - 588)
 INTEGRAL Likelihood = -0.32 Transmembrane 665 - 681 (665 - 681)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4598(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF16724 GB:AF141644 putative integral membrane protein
 [Lactococcus lactis]
 Identities = 109/195 (55%), Positives = 156/195 (79%), Gaps = 4/195 (2%)
 Query: 466 KMFNNGLASLKKSGEYDKLVKKYLSTASTSSNDKAAKPVDESTILGLISNNYKQLLSGIG 525
 +MFNNGLA+L+ +GEYDK++ KYL++ T + +AK E+T G++ NN++Q+ G+
 Sbjct: 1 EMFNNGLANLRANGEYDKIIDKYLAS-DTKTIQSSAK---ENTFFGILQNNWEQIGRGLL 56
 Query: 526 TTLSLTILISFAIAMVIGIIFGMMSPSNTLRTISMIFVDIVRGIPLMIVAFAFIFWGIPN 585
 TL L ++SF +AM++GIIFG+ SV+PS LRTI+ I+VD+ R IPL+++ FIF+GIPN
 Sbjct: 57 VTLELAVLSFILAMIVGIIFGLFSVAPSKILRTIARIYVDLNRSLPLLVLTIFIFYGIPN 116
 Query: 586 LIESITGHQSPINDFVAATIALSLNGGAYIAEIVRGGIEAVPSGQMEASRSLGISYGTKM 645
 L++ ITGHQSP+N+F A IAL+LN AYIAEIVR G++AVPSGQMEASRSLG++Y +M
 Sbjct: 117 LLQIITGHQSPPLNEFTAGVIALTLNSSAYIAEIVRSGVQAVPSGQMEASRSLGVTYLTSM 176
 Query: 646 QKVILPQAVRLMLPN 660
 +KVILEQA+++ +P+
 Sbjct: 177 RKVILPQAIKITIPS 191

There is also homology to SEQ ID 1198.

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9071> which encodes amino acid sequence <SEQ ID 9072>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> May be a lipoprotein
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1448-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 80.8 bits (196), Expect = 2e-17
 Identities = 64/233 (27%), Positives = 113/233 (48%), Gaps = 13/233 (5%)

Query: 34 IKKTRKLVAVSPDYAPFEFKALVNGKDTIVGADVQLAQAIADLVDLELSPMSFDNVL 93
 +K + K+V S +APFE++ NGK G D++L + IA + L++S FD L
 Sbjct: 268 VKPSYKIVSDSS--FAPFEYQ---NGKGKYGTFDMELIKKIAKQGGFKLDISNPGFDAAL 322

Query: 94 SSLQTGKADLAISGISHTKERAKVYDFSIPYYQAENAIVMRASDAKVTKNISDLNGKKVA 153
 +++Q+G+AD I+G + T+ R K++DFS PYY +++++ K+ DL GK V
 Sbjct: 323 NAVQSGQADGVIAGATTTEARQKIFDFSDFPY--TSSVILAVKKGSNVKSYQDLKGKTVG 380

Query: 154 AQKGSIEEGLVKIQLPKANLISLTAMGEA---INELKAGQVYAVTLEAFVAAGFLAQHKD 210
 A+ G+ + K N + A EA + + +G + A+ + V A + Q +
 Sbjct: 381 AKNGTASYTWLSDHADKYN-YHVKAFFDEASTMYDSMNSGSIDALMDDEAVLAYAINQGRK 439

Query: 211 LALAPFSLKTSDDAKAVALPKNSGDLTKAVNKVIAKLDEQERYKSFIETIA 263
 P + S GD + +L K N +A L + Y + + ++
 Sbjct: 440 FE-TPIKGEKS-GDIGFAVKKGANPELIKMFNNGLASLKKSGEYDKLVKKYLS 490
 Score = 74.5 bits (180), Expect = 1e-15
 Identities = 59/215 (27%), Positives = 102/215 (47%), Gaps = 12/215 (5%)

Query: 48 YAPFEFKALVNGKDTIVGADVQLAQAIADLVDLELSPMSFDNVLSLQTGKADLAISG 107
 YAPFEFK + T G DV + +A ++ ++ FD ++++Q+G+AD ++G
 Sbjct: 36 YAPFEFK---DSDQTYKGIDVDIVNEVAKRAGWNVNMTYPGFDAAVNAVQSGQADALMAG 92

Query: 108 ISHTKERAKVYDFSIPYYQAENAIVMRASDAKVTKNISDLNGKKVAAQKGSIEEGLVKIQ 167
 + T+ R KV++FS YY + I+ ++ KVT N L GK V + G+ + ++
 Sbjct: 93 TTVTEARKKVFNFSDTYYDT-SVILYTKNNNKVT-NYKQLKGKVVGVKNGTAAQSFLEEN 150

Query: 168 LPKANLISLTAMGEAI--NELKAGQVYAVTLEAFVAAGFLAQHKDLALAPFSLKTSDDGA 225
 K T + N L +G +YA + PV + Q K A+ +++ +
 Sbjct: 151 KSKYGYKVKTFDTSDLMNNSLDSGSIYAAMDDQFVVQFAINQGGKAYAI---NMRGEAVGS 207

Query: 226 KAVALPKNSG--DLTKAVNKVIAKLDEQERYKSFI 258
 A A+ K SG +L K N A++ Y +
 Sbjct: 208 FAFVKKGSCHDNLIKEFNATAFAQMKSDGTYNIDIM 242

SEQ ID 4024 (GBS154) was expressed in *E.coli* as a His-fusion product. The purified protein is shown in Figure 199, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1316

A DNA sequence (GBSx1396) was identified in *S.galactiae* <SEQ ID 4025> which encodes the amino acid sequence <SEQ ID 4026>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein (glnQ). Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4183(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB90561 GB:AE001058 glutamine ABC transporter, ATP-binding

-1449-

protein (glnQ) [Archaeoglobus fulgidus]
Identities = 147/240 (61%), Positives = 192/240 (79%)

5 Query: 5 KIDVQDLHKSQNEVLKQIDAKFYEGDVVCIIGPSGSGKSTFLRTLNLLESITSGKVVV 64
++++ DLHK +G+ EVLKG+ K +G+VV IIGPSGSGKST LR +N LE TSGK+++
Sbjct: 3 QLEIIDLHKRFGELEVLKGVMTMKVEKGEVVVVIIGPSGSGKSTLLRCINRLEPTSGKILL 62

10 Query: 65 DGFELSNPKTDIDKARENIGMVQHFNLFPHMSVLENITFAPIELGKESKEAAEKHGMEL 124
DG +++N K DI+K R+ IG+VFQ FNLFPH++ L+N+T API++ K SK AE+ GM L
Sbjct: 63 DGVDITNSKIDINKVRQIRIGIVFQQFNLFPHTALQNVTLAPIKIKKMSKREAEELGMRL 122

15 Query: 125 LEKVGLADKANAKPDSLGGQKQORVAIARSLAMNPDILLFDEPTSALDPEMVGDLNVMK 184
LEKVGL DKA+ P LSGGQ+QORVAIAR+LAMNP+++LFDE TSALDPE+V +VL+VMK
Sbjct: 123 LEKVGLADKADYYPQLSGGQKQORVAIARALAMNPEVMLFDEVTSALDPELVKEVLDVMK 182

Query: 185 DLAEQGMTMLIVTHEMGFARQVANRVIFTDGGRFLEDGTPEQIFDTPQHPRLQDFLNKVL 244
LA GMTM++VTHEMGFAR+V +RVIF DGG +E+G PEQIF P+H R + FL+ +L
Sbjct: 183 QLARDGMTMVVVTHEMGFAREVGDRVIFMDGGVIVEEGKPEQIFSNPKHERTRKFLSMIL 242

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4027> which encodes the amino acid sequence <SEQ ID 4028>. Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4149(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the databases:

>GP:BAB05180 GB:AP001512 ABC transporter (substrate-binding protein)
[Bacillus halodurans]
Identities = 79/227 (34%), Positives = 126/227 (54%), Gaps = 10/227 (4%)

35 Query: 35 KKTRKLVVAVSPDYAPFEFKALVNGKDTIVGADVQLAQAIADDELVDLELSPMSFDNVL 94
+K LV+ S DY P+E + G+ IVG DV +A+ I EL +L++ M F+ ++
Sbjct: 48 EKKSVLVMGTSADYPPYESVDVTG--IVGFDVDIAEYITSELGYELKIQDMDENGIIIP 105

40 Query: 95 SLQTKADLAISGISHTKERAKVYDFSIPYYQAEINAIVMRASDAKVTKNISDLNGKKVAA 154
+LQ G+ D A+SG++ T+ER K DFS YY A+N +V + D ++ DL GK V
Sbjct: 106 ALQAGRVDFALSGMTPTBERKKSVDVSDVYDAQNLVVFKEDG--LSSVEDLAGKTVGV 163

45 Query: 155 QKGSII-EEGLVKIQ--LPKANLISLTAMGEAINELKAGQVYAVTLEAPVAAGFLAQHKDL 211
Q SI EE V++Q L + + + E + EL AG+V A+ +E VAAG L +
Sbjct: 164 QLASIQEEAAVELQEELDGLTIETNRNRPVQLVQELLAGRVDALIIEDTVAAGHLEANP-- 221

Query: 212 ALAPFSLKTSDDGDAKAVLPKNSGDLTKAVNKVIAKLDEQERYKSFI 258
L F++++ A+A PK+S +LT+ N+ + ++ E + I
Sbjct: 222 GLVRFAIESEGETGSAIAFPKDS-ELTEPFNEKLQEMMEDGTMEELI 267

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 223/246 (90%), Positives = 238/246 (96%)

55 Query: 1 MAELKIDVQDLHKSQNEVLKQIDAKFYEGDVVCIIGPSGSGKSTFLRTLNLLESITSG 60
M ELKIDVQDLHKSQNEVLKQIDAKFYEGDVVCIIGPSGSGKSTFLRTLNLLE+ITSG
Sbjct: 1 MTELKIDVQDLHKSQNEVLKQIDAKFYEGDVVCIIGPSGSGKSTFLRTLNLLETITSG 60

60 Query: 61 KVVVDGFELSNPKTDIDKARENIGMVQHFNLFPHMSVLENITFAPIELGKESKEAAEKH 120
KV+VDGFELS+PKT+IDKARENIGMVQHFNLFPHM+VLENI FAP+ELGKESKE A+KH
Sbjct: 61 KVMVDGFELSDPKTNIDKARENIGMVQHFNLFPHMTVLENTIFAPVELGKESKEVAKH 120

Query: 121 GMELLEKVGLADKANAKPDSLGGQKQORVAIARSLAMNPDILLFDEPTSALDPEMVGDLV 180
GM LLEKVGL+DKA+A P LSGGQKQORVAIARSLAMNPDI+LFDEPTSALDPEMVGDLV
Sbjct: 121 GMALLEKVGLSDKADAFPGSLGGQKQORVAIARSLAMNPDIIMLFDEPTSALDPEMVGDLV 180

-1450-

Query: 181 NVMKDIAEQGMTMLIVTHEMGFARQVANRVIFTDGGRFLEDGTPEQIFDTPQHPRQLQDFL 240
 NVMKDIAEQGMTMLIVTHEMGFARQVANRVIFTDGG+FLEDGTPE+IFD P+HPRL +FL
 Sbjct: 181 NVMKDIAEQGMTMLIVTHEMGFARQVANRVIFTDGGQFLEDGTPEEIFDHPKHPRLIEFL 240

Query: 241 NKVLNV 246
 +KVLNV
 Sbjct: 241 DKVLNV 246

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1317

A DNA sequence (GBSx1397) was identified in *S.agalactiae* <SEQ ID 4029> which encodes the amino acid sequence <SEQ ID 4030>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.2311(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4031> which encodes the amino acid sequence <SEQ ID 4032>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.2702(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 45/51 (88%), Positives = 49/51 (95%)

Query: 1 MGDKPISFRDKDGNFVSAADVWNAEKLEELFNTLNPNRKRLRLEREKLAKKEK 51
 MGDKPISF+DKDGNFVSAADVWNAEKLEELFN LNPNR+LRLEREKL K++
 Sbjct: 11 MGDKPISFKDKDGNFVSAADVWNAEKLEELFNLLNPNRRLRLEREKLKKDE 61

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1318

A DNA sequence (GBSx1398) was identified in *S.agalactiae* <SEQ ID 4033> which encodes the amino acid sequence <SEQ ID 4034>. This protein is predicted to be spo0b-associated GTP-binding protein (obg). Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.2967(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

-1451-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14752 GB:Z99118 GTPase activity [Bacillus subtilis]

Identities = 297/435 (68%), Positives = 345/435 (79%), Gaps = 7/435 (1%)

5 Query: 3 MFLDTAKISVKAGRGDGMVAFRREKYVPNGGPWGGDGGKGGSVIFKVN EGLRTLMDFRY 62
MF+D K+ VK G GG+GMVAFRREKYVP GGP GGDGGKGG V+F+V+EGLRTLMDFRY
Sbjct: 1 MFVDQVKVYVKGDDGGMVAFRREKYVPKGGPAGDGGKGGDVVFEVDEGLRTLMDFRY 60

10 Query: 63 NRNFKAKAGEKGMTKGMHGRGAEDLIVSLPPGTTVRDATTGKVITDLVEHDQEFVVARGG 122
++FKA GE GM+K HGR A+D+++ +PPGT V D T +VI DL EH Q V+ARGG
Sbjct: 61 KKHFKAIRGEHGM SKNQHGRNADDMVIVPPGTVVTD DDTKQVIADL TEHQRAVIARGG 120

15 Query: 123 RGGGRGNIRFATPRNPAP EIAENGEPGEERELQLELKILADVGLVGFPSVGKSTLLSVVSA 182
RGGGRGN RFATP NPAP+++ENGEPG+ER + LELK+LADVGLVGFPSVGKSTLLSVV+
Sbjct: 121 RGGGRGNSRFATPANPAPQLSENGEPGKERYIVLELKV LADVGLVGFPSVGKSTLLSVVSS 180

20 Query: 183 AKPKIGAYHFTTIVPNLGMVRTKSGDSFAMADLPGLIEGASQGVGLGTQFLRHIERTRVI 242
AKPKI YHFTT+VPNLGMV T G SF MADLPGLIEGA QGVGLG QFLRHIERTRVI
Sbjct: 181 AKPKIADYHFTTIVPNLGMVETDDGRSFVMADLPGLIEGAHQGVGLGHQFLRHIERTRVI 240

25 Query: 243 LHVIDMSASEGRDPYDDYVSINNELETYNLRLMERPQIIIVANKMDMPDSEENLA AFKEKL 302
+HVIDMS EGRDPYDDY++IN EL YNLRL ERPQIIIVANKMDMP++ ENL AFKEKL
Sbjct: 241 VHVIDMSGLEGRDPYDDYLTINQELSEYNLRLTERPQIIIVANKMDMPEAAENLEAFKEKL 300

30 Query: 303 AANYDEFDDMPMIFPISSLAHQGLENLMDATAELLANT EEFLLYDETD MQEDEAYYGFNE 362
DD P +FPIS++ +GL L+ A L NT EF LYDE ++ ++ Y
Sbjct: 301 T-----DDYP-VFPISAVTREGLRELLFEVANQLENTPEFFLYDEBELTQNRVMYTMEN 353

35 Query: 363 DERPF EITRDDDATWVLYGDKLEKLFVMTNMERDESIMK FARQLRGMGVDEALRERGAKD 422
+E PF ITRD D +VL GD LE+LF MT+ RDES+ +FARQ+RGMGVDEALRERGAKD
Sbjct: 354 EEVFPNITRDPDGVFVLSGDSLERLFPKMTDFSRDES VKRFARQMRGMGVDEALRERGAKD 413

Query: 423 GDIVRIGNFEFEFVD 437
GDI+R+ FEFEF+D
Sbjct: 414 GDIIIRLLEFEFEFID 428

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4035> which encodes the amino acid sequence <SEQ ID 4036>. Analysis of this protein sequence reveals the following:

40 Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2588(Affirmative) < succ>
45 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 394/437 (90%), Positives = 421/437 (96%)

50 Query: 1 MS MFLDTAKISVKAGRGDGMVAFRREKYVPNGGPWGGDGGKGGSVIFKVN EGLRTLMDF 60
MS MFLDTAKISV+AGRGDGMVAFRREKYVPNGGPWGGDGGKGGSVIF+V+EGLRTLMDF
Sbjct: 1 MS MFLDTAKISVQAGRGDGMVAFRREKYVPNGGPWGGDGGKGGSVIFRVDEGLRTLMDF 60

55 Query: 61 RYNRNFKAKAGEKGMTKGMHGRGAEDLIVSLPPGTTVRDATTGKVITDLVEHDQEFVVAR 120
RYNR FKAK+GEKGMTKGMHGRGAEDLIV +P GTTVRDA TGKVITDLVEH QE V+A+
Sbjct: 61 RYNRNFKAKSGEKGMTKGMHGRGAEDLIVFPQGTTVRDAETGKVITDLVEHQEVVIAK 120

60 Query: 121 GGRGGRGNIRFATPRNPAP EIAENGEPGEERELQLELKILADVGLVGFPSVGKSTLLSVV 180
GGRGGRGNIRFATPRNPAP EIAENGEPGEER+L+LELKILADVGLVGFPSVGKSTLLSVV
Sbjct: 121 GGRGGRGNIRFATPRNPAP EIAENGEPGEERQLELELKILADVGLVGFPSVGKSTLLSVV 180

Query: 181 SAAKPKIGAYHFTTIVPNLGMVRTKSGDSFAMADLPGLIEGASQGVGLGTQFLRHIERTR 240
S+AKPKIGAYHFTTIVPNLGMVRTKSGDSFAMADLPGLIEGASQGVGLGTQFLRHIERTR

-1452-

Sbjct: 181 SSAKPKIGAYHFTTIVPNLGMVVRTKSGDSFAMADLPGLIEGASQGVGLGTQFLRHIERTR 240

Query: 241 VILHVIDMSASEGRDPYDDYVSINNELETYNRLMERPQIIIVANKMDMPDSEENLAAPKE 300
 VILHVIDMSASEGRDPY+DYVSINNELETYNRLMERPQIIIVANKMD+P+++ENL AFK+

5 Sbjct: 241 VILHVIDMSASEGRDPYEDYVSINNELETYNRLMERPQIIIVANKMDIPEAQENLKAFKK 300

Query: 301 KLAANYDEFDDMPMIFPISSLAHQGLENLMDATAELLANTEEFLLYDETDMOEDEAYYGF 360
 KLAA YDEFDD+PMIFPISSLAHQGLENL++ATAELLA T+EFLLYDE+D+ ++EAYYGF

10 Sbjct: 301 KLAAYDEFDDLPMIFPISSLAHQGLENLLEATAELLAKTDEFLLYDES DLVDEEAYYGF 360

Query: 361 NEDERPFEITRDDDATTWVLYGDKLEKLFVMTINMERDESIMKFARQLRGMGVDEALRERGA 420
 E E+ FEITRDDDATTWV G+KLE+LFVMTINMERDESIMKFARQLRGMGVDEALRERGA

Sbjct: 361 AETEKDFEITRDDDATTWVLSGEKLERLFVMTINMERDESIMKFARQLRGMGVDEALRERGA 420

15 Query: 421 KDGDIVRIGNFEFEFVD 437
 KDGD VRIG FEFEFVD

Sbjct: 421 KDGDVPRIGKFEFEFVD 437

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1319

A DNA sequence (GBSx1399) was identified in *S.agalactiae* <SEQ ID 4037> which encodes the amino acid sequence <SEQ ID 4038>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4039> which encodes the amino acid sequence <SEQ ID 4040>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 40 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 30/42 (71%), Positives = 37/42 (87%)

45 Query: 1 MAFGDNGQRKKTGFEEKLTILFVVILMVLVTVGGLVFGAISAIM 42
 +AFG+NG RKKT FEK+T+FVVILMVLVTVGGL+ A+S +M

Sbjct: 1 VAFGENGPRKKTTFEKTMTFVVILMVLVTVGGLIASALSVLM 42

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1320

A DNA sequence (GBSx1401) was identified in *S.agalactiae* <SEQ ID 4041> which encodes the amino acid sequence <SEQ ID 4042>. Analysis of this protein sequence reveals the following:

-1453-

Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2484(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAD28348 GB:AF102860 aminopeptidase PepS [Streptococcus hermophilus]
 Identities = 247/413 (59%), Positives = 313/413 (74%)

Query: 1 MVLQDFDNLKKYAQLIISKGLNVQKGHITLALTIDVEQVHLARLLTEAAYEKGASEVIVD 60
 MVL +F L+KYA+L+++ G+NVQ GHT+AL+IDVEQ LA LL + AY GA+EVIV

15 Sbjct: 1 MVLPNFKENLEKYAKLLVTNGINVQPGHTVALSIDVEQAEHLAHLVKEAYALGAAEVIVQ 60

Query: 61 YTDDFITRQRLHASDEVLTNPVQYTVDKSLALLNKKASRLVVKSSNPNAFATVDPKRLS 120
 ++DD I R+R LHA + VP Y + LL KKASRL V+SS+P+AF V P+RLS

20 Sbjct: 61 WSDDTINRERFLHAEMNRIEEVPAYKKAEMEYLLKKASRLGVRSSDPDAFNGVAPERLS 120

Query: 121 ETTRATAIALEEQSRATQANKVSWNVAAAAGREWAALVFPPELKTSDQQVDALWDTIFKLN 180
 +A A + A Q+NKVSU VAAAAG+EWA VFP + ++ VD LW+ IFK

25 Sbjct: 121 AHAKAIGAFAKPMQVATQSNKVSUWVAAAAGKEWAKVFPNASSDEEAVDLLWNQIFKTC 180

Query: 181 RIYEDDPIAAWDAHEAKLLEKATRLNQEQFDALHYTAPGTDLTGLMPKNHIWEAAGSLNA 240
 R+YE DP+ AW H +L KA LN+ QF ALHYTAPGTDLTGL+PKNH+WE+AG++NA

30 Sbjct: 181 RVYEKDPVRAWKEHADRLDAKARILNEAQFSALHYTAPGTDLTGLPKNHVWESAGAINA 240

Query: 241 QGETFIANMPTEEIFSAPDYRRADGYVTSTKPLSYAGVIIENMTFTFKDGKIINVTAKEG 300
 QGE+F+ NMPTEE+F+APD+RRA GYV+STKPLSY G IIE + TFKDG+I+++TA++G

35 Sbjct: 241 QGSEFLPNMPTTEEVFTAPDFRRAYGYVSTKPLSYNGNIIIEGKIVTFKDGGEIVDITADQG 300

Query: 301 QETVQRLIEENDGARSIGEVAVLPHKTPISLSGLIFFNTLFDENASNHLAIGTAYAFNVE 360
 ++ ++ L+ N+GAR+LGE ALVP +PIS SG+ FFNTLFDENASNHLAIG AYA +VE

40 Sbjct: 301 EKVMKNLVFNNGARALGECALVPDSSPISQSGITFFNTLFDENASNHLAIGAYATSVE 360

Query: 361 GGTEMTSQELDEAGLNRSSTHVDFMIGSEQMDIDGIRADGTAVPIFRNGEWAI 413
 GG +MT +EL AGLNRS HVDF+IGS QM+IDGI DG+ VPIFRNG+W I

45 Sbjct: 361 GGADMTEEELKAAGLNRSVDHVDFIIGSNQMNIDGIHHDGSRVPIFRNGDWVI 413

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1321

45 A DNA sequence (GBSx1403) was identified in *S.agalactiae* <SEQ ID 4045> which encodes the amino acid sequence <SEQ ID 4046>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have a cleavable N-term signal seq.

50 INTEGRAL Likelihood = -7.91 Transmembrane 661 - 677 (657 - 680)

----- Final Results -----

55 bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8787> which encodes amino acid sequence <SEQ ID 8788> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7

McG: Discrim Score: 6.47

-1454-

GvH: Signal Score (-7.5): 1.01
 Possible site: 29
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 1 value: -7.91 threshold: 0.0
 5 INTEGRAL Likelihood = -7.91 Transmembrane 658 - 673 (657 - 680)
 PERIPHERAL Likelihood = 4.35 555
 modified ALOM score: 2.08

*** Reasoning Step: 3

10 ----- Final Results -----
 bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 15 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 647-651

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAF09821 GB:AE001885 6-aminohexanoate-cyclic-dimer hydrolase
 [Deinococcus radiodurans]
 Identities = 150/497 (30%), Positives = 233/497 (46%), Gaps = 32/497 (6%)

Query: 110 LTEETYKQKDGQDLANMVRSGQVTSEELVNMAYDIIAKENPSLNAVITTRQEAIEBARK 169
 LT Y + D DLA + R G++++E++ A N +LNAV+ + + +AR
 25 Sbjct: 45 LTFAEYDRDLDAQLFRRGELSAEDMCTAATHRAQVVNVALNAVVPYLYDQGLAQARA 104

Query: 170 L-----KDTNQPFLLGVPLLVKGLGHSIKGGETNGLIYADGKISTFDSSYVKKYKDLG 222
 + PF GVP LVK G + G G +I +D V++++ G
 30 Sbjct: 105 TDAARARGEQATGPFAGVPFLVKDFGSRLAGVPHTGGTRAYRDQIFEWDELVRRWQAAG 164

Query: 223 FIILGQINFFEYGWRNITDSKLYGLTHNPWDLAHNAGSSGGSAAAIAAGMTPIASGSDA 282
 + LG+IN PE+ +T+ +L+G T NPWDL GGSSGGSAA+A+A+G+ P+A D
 35 Sbjct: 165 LLPLGKTINTPEFALMGVTEPELHGFTRNPNWDLGRTPGGSSGGSASAVAAGIVPLAGAGDG 224

Query: 283 GGSIRIPSSWTGLVGLKPTRGLV---SNEKPDSYSTAVHFPPLTKSSRDAETLLTYLKKS 339
 GGSIRIP+S GL GLKP+RG V AV LT+S RD+ LL + D
 40 Sbjct: 225 GGSIRIPASCCGLFGLKPSRGRVPCGDGVGEWPQGAAVEHVLTSSVRDAAALLDLEQGP 284

Query: 340 QTLVSV-----NDLKSPLIAYTLKSPMGTEVSQDAKNAIMDNVTFLRKQGF 386
 + L I ++ P+G V + A+ L G +
 45 Sbjct: 285 AGAALFLPSPERPYSEEVGREPGRLRIGFSTAHPGRSVHPECVAAVQGAARLLES LGHE 344

Query: 387 VTEIDLPIIDGRALMRDYSTLAIGMGGAFFSTIEKDLKKHGFTKEDVDPIWAVHVIYQNSD 446
 V E+ LP DG AL + + L G GA +D DV+ +TW + + ++
 50 Sbjct: 345 VEEVALPWDGPALAQAFMLLYFGETGASLAALRDTLGRPARASDVEAVTWLLGQLGRSYS 404

Query: 447 KAELKKSIMEAQKHMDYRKAMEKLHKQFFIFLSPTTASLAPLNTDPY----VTEEDKRA 502
 A+ A+ + + +AM + H+ + + L+P A+ PL V RA
 55 Sbjct: 405 AAD----FAAARASWNVHARAMGRFHQNYDLLLTPLVAT-PPLQIGELQPRGVQAALLRA 459

Query: 503 IYNMENLSQEERIALFNROWEFMLRRTPTQIANMTGLPAISIPTYLSESGLPITGLMLMA 562
 M+ R + +L + P+TQ+AN+TG PA+S+P + + GLP+G +A
 60 Sbjct: 460 AQQMDVSGLLRRSGQVDALATDILEKMPYQLANLTGQPAMSVPLHWTADGLPVGVQFVA 519

Query: 563 GANYDMVLIKFAFFFEK 579
 + VL++ A E+
 65 Sbjct: 520 PLAREDVLLRLAGQLEQ 536

There is also homology to SEQ ID 4048.

60 SEQ ID 8788 (GBS173) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 5; MW 96.8kDa).

The GBS173-GST fusion product was purified (Figure 116A; see also Figure 201, lane 7) and used to immunise mice (lane 1+2 product; 15µg/mouse). The resulting antiserum was used for Western blot, FACS,

-1455-

and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1322

A DNA sequence (GBSx1404) was identified in *S.agalactiae* <SEQ ID 4049> which encodes the amino acid sequence <SEQ ID 4050>. This protein is predicted to be ribosomal large subunit pseudouridine synthase B (rsuA). Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3674(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB06992 GB:AP001518 16S pseudouridylate synthase [Bacillus halodurans]
Identities = 110/236 (46%), Positives = 149/236 (62%), Gaps = 4/236 (1%)

20 Query: 1 MRLDKFLVECGLSRTQVKLILKKKQISVNGNSETSPKVQVDEYRDEIKYNGTLVSYEKF 60
      MR+DKFL G GSR VK +LK + V G P V+ + I G V Y+ +
      Sbjct: 1 MRIDKFLANMGFGSRKDVKKLKTGAVRVQGQPIKDPSTHVEPESESITVYGEFVEYKPY 60

25 Query: 61 VYYMLHKPKGVISATDDPSHKTVDLLDKTARDKAVFPVGRLDIDTTGLLLLTNNGELAH 120
      VY M++KPKGVI AT+D H+TV+DLL + R PVGRLD DT GLLL+TN+G+ H
      Sbjct: 61 VYLMNPKPKGVICATEDLEHETVIDLLGEEERHYEPSPVGRLDKDTVGLLLITNDGKFNH 120

30 Query: 121 KMLSPKKHVDKCYEVKISGIMTEDDILAFDKGIILKD-FTCLPALLEIVEVNQVKQSLV 179
      ++SPK HV K Y + G +TE+D+ AF G++L D + FA L I+E +S +
      Sbjct: 121 WLMSPKHHVPKTYRALVEGHVTEEDVGAFSHGVVLDGYYVTKPATLHILEAG---ARSHI 177

35 Query: 180 KITIKEGKFHQVKRMVAACGKEVLELKRMRGNLQLDKQLESGQWRRLTIKEIEKL 235
      ++ + EGKFHQVKRM A GK VLEL+R+++GNL LD +L G++R LT +EI L
      Sbjct: 178 ELILTEGKFHQVKRMFQAVGKRVLELERIKIGNLLLDPELARGEYRELTKEEIALI 233

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4051> which encodes the amino acid sequence <SEQ ID 4052>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0152(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAF09821 GB:AE001885 6-aminohexanoate-cyclic-dimer hydrolase
[Deinococcus radiodurans]
50 Identities = 177/485 (36%), Positives = 259/485 (52%), Gaps = 13/485 (2%)

Query: 5 DATAMAIAVQTGQTTPLELVTOAIYKAKKLNPTLNAITSERFEAALEAKQDFSGL--- 61
      DA +A + G+ + ++ T AI++A+ +N LNA+ ++ L +A+ D +
      Sbjct: 54 DALDLAQLFRRGELSAEDMCTAAIHRAQVVNVALNAVYPLYDQGLAQARATDAARARGE 113

55 Query: 62 ----PFAGVPLFLKDLGQELKGSSTSGSRLFKKEYQATKTDLFVKRLEALGFIIIGRSNT 117
      PFAGVP +KD G L G T G+R +++ D V+R +A G + LG++NT

```

-1456-

Sbjct: 114 QATGPFAGVPFLVKDFGSRLAGVPHTGGTRAYRDQIPEWDDLVRRWQAAGLLPLGKTINT 173

Query: 118 PEFQFKNISDSSLHGPNLPRDNTRNAGSSGGAAALVSSGISALATASDGGGSIRIPAS 177
 PEF +++ LHGP P D R GGSSGG+A+ V++GI LA A DGGGSIRIPAS

5 Sbjct: 174 PEFALMGVTEPELHGPTRNPNWDLGRTPGGSSGGSASAVAAGIVPLAGAGDGGGSIRIPAS 233

Query: 178 FNGLIGLKPSRGRMPVPGFSYRSWQASVHFALTKSVRDTRNLLYYLQMEQMESPFPLAT 237
 GL GLKPSRGR+P G G WQGA+V LT+SVRD+ LL Q + L +

10 Sbjct: 234 CCGFLGFLKPSRGRVPCGDGVGEFWQGA+VEHVLTRSVRDSAAALLDLEQGPDAAGALFLPS 293

Query: 238 LTKDSIYQSLQRP--LTIAFYQRLSDGSPVSLDTAKALRQAVTWLREQGHQVLEEEFPV 295
 + + + P L I F G V + A++ A L GH++ E+ P

Sbjct: 294 PERPYSEEVGREPRLRIGFSTAHLGRSVHPECVAAVQGAARLLES LGHEVEEV-ALPW 352

15 Query: 296 NMTEVIRHYIMNSVETAAMFADIEDTFGRPMTKDDMETMTWAIYQSGKDIPAWRYSQVL 355
 + + + + + ET A A + DT GRP D+E +TW + Q G+ A ++

Sbjct: 353 DGPALAAQAFMLYFGETGASLAALRDTLGRPARASDVEAVTWLLGQLGRSYSAADFAAAR 412

20 Query: 356 QKWDYTSATMASFHETYDLLLTFTTNTPAKPHGELVP---DSKLMANLAQAEIFSSEEQF 412
 W+ ++ M FH+ YDLLLT TP + GEL P + L+ Q ++ +

Sbjct: 413 ASWNVHARAMGRFHQNYDLLLTFPVLATPPLQIGELQPRGVQAALLRAAQMDVSGLLRRS 472

Query: 413 NLVETMFGKSLAINPYTALPNLTGQPAISLPTYETKEGLSMGIQLIAAKGREDLLLGIAE 472
 V+ + L PYT L NLITGQPA+S+P + T +GL +G+Q +A RED+LL +A

25 Sbjct: 473 GQVDALATDILEKMPYTQLANLTGQPAMSVPLHWTADGLPVGVQFVAPLAREDVLLRLAG 532

Query: 473 QFEAA 477
 Q E A

30 Sbjct: 533 QLEQA 537

An alignment of the GAS and GBS proteins is shown below.

Identities = 151/240 (62%), Positives = 183/240 (75%)

Query: 1 MRLDKFLVECGLSRTQVKLILKKKQISVNGNSETSPKVQVDEYRDEIKYNGTLVSYEKF 60
 MRLDKFLV G+G+R+QVKL+LKKK I VN ETS K +DEY+D + Y GT + YE F

35 Sbjct: 2 MRLDKFLVATGVGTRSQVKLILKKKKAIFVNQKVETSAKAHIDEYKDLVTYQGTPLVYESF 61

Query: 61 VYYMLHKPKGVISATDDPSHKTVDLLDKTARDKAVFPVGRLDIDTTGLLLLTNNGLAH 120
 VYY+L+KP G +SAT D TV++LLD TAR KAVFPVGRLD DT GLLLLTNNG+LAH

40 Sbjct: 62 VYYLLNKPSCGYVSATQDRQATVMELDDTARQKAVFPVGRLDKDTRGLLLLTNNGLAH 121

Query: 121 KMLSPKKHVDKCYEVKISGIMTEDDILAFDKGIILKDFTCPLALLEIVEVNOVKQSLVK 180
 +LSPKKHV K Y K++GIMTE D F +GI LKD CLPA LE++ + ++ SLVK

45 Sbjct: 122 DLLSPKKHVTKEYLAKVAGIMTEADKDYFARGISLKDHCPLAHLEVLASDLQQTSLVK 181

Query: 181 ITIIEGKFHQVQRMVAACGKEVLELRLRMGNLQLDKQLESGQWRRLLTIKEIEKLEKYM 240
 ITI+EGKFHQVQRMVAACGKEVL+L+RL MG L+LD L G++RRLT +E++ L Y Q

Sbjct: 182 ITIQEGKFHQVQRMVAACGKEVLDLQRLSMGPLKLDPSLAEGEFRRLLTPEELQSLAPYCQ 241

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1323

A DNA sequence (GBSx1405) was identified in *S.agalactiae* <SEQ ID 4053> which encodes the amino acid sequence <SEQ ID 4054>. Analysis of this protein sequence reveals the following:

55 Possible site: 46
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

60 bacterial cytoplasm --- Certainty=0.2811(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1457-

A related GBS nucleic acid sequence <SEQ ID 10007> which encodes amino acid sequence <SEQ ID 10008> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAA57350 GB:J04483 reductase [Leishmania major]
    Identities = 129/277 (46%), Positives = 167/277 (59%), Gaps = 3/277 (1%)

    Query: 26  TLSNTLNIPKIGFGTWQLTEGEEAYKAVTHALKVGYTHIDTAQIYGNESVSGRAIRDSGL 85
              TLSN + +P+ G G WQ  GE  AV  AL  GY HIDTA IY NE SVG +R SG+
    10  Sbjct: 10  TLSNGVKMPQFGLGVWQSPAGEVTENAVNWALCAGYRHIDTAAYKNEESVSGAGLRASGV 69

    Query: 86  ARESIFLTITKIWNKHDYHLAKASIDESLQKLGVDYIDLLLIHWPNPKALRENDKAGN 145
              RE +F+ITK+WN + Y  A+ +ES QKLGVDYIDL LIHWP K + + K
    15  Sbjct: 70  PREDVFITTKLWNTTEQGYESTLAAFEESRQKLGVDYIDLYLIHWPRGKDILSKEGKXY-- 127

    Query: 146  AGTWKAMEEAYKEGKVKVKAIGVSNFMKHHLEALFETAEIKPMVNQIILAPGCAQEDLVRFC 205
              +W+A E+ YKE KV+AIGVSNF HHLE + + PMVNQ+ L P  Q DL FC
    20  Sbjct: 128  LDSWRAFEQLYKEKKVRAIGVSNFHHLEEDVLAMCTVTPMVNQVELHPLNNQADLRAFC 187

    Query: 206  KGNIDILLEAYSPFGTGAIFENESIKAIKAEKYGKSVQVALRWSLDNGFLPLPKSATPKNI 265
              I +EA+SP G G + N + AI KY K+ AQV LRW++ + +PKS + I
    25  Sbjct: 188  DAKQIKVEAWSPLGQGLLSNFILSAIGAKYNKTAQVILRWNIQKNLITIPKSVHRERI 247

    Query: 266  EANLDIFDFQLNEDDIATLIQLDSGIK-PKDPDNVSF 301
              E N DIFDF+L +D+ ++ L++ + DPD F
    30  Sbjct: 248  EENADIFDFELGAEDVMSIDALNTNSRYGPDPEAQF 284
  
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 779> which encodes the amino acid sequence <SEQ ID 780>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 27
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.0980(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    35  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  
```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 155/282 (54%), Positives = 204/282 (71%), Gaps = 2/282 (0%)

    40  Query: 20  IVMETYTLNIPKIGFGTWQLTEGEEAYKAVTHALKVGYTHIDTAQIYGNESVGRA 79
              +++ T +++ IP +GFGT+Q +GEEAY++ A+K GY HIDTA IY NE SVGRA
    Sbjct: 1  VMVTTVKMTSGYEIPVLGFGTYQAADGEEAYQSTLAAIKAGYRHIDTAAYKNEESVGRA 60

    45  Query: 80  IRDSGLARESIFLTITKIWNKHDYHLAKASIDESLQKLGVDYIDLLLIHWPNPKALREND 139
              I+DSG+ RE +F+ITK+WND H Y AK ++ SL +LG+DY+DL LIHWPNPKALR +
    Sbjct: 61  IKDSGVLRRDLFITTKLWNDHAHSYEGAKDALAASLDRLGLDYVDLYLIHWPNPKALR--N 118

    Query: 140  AWKAGNAGTWKAMEEAYKEGKVKVKAIGVSNFMKHHLEALFETAEIKPMVNQIILAPGCAQE 199
              WK NA W+ MEEA + G +K+IGVSNFM HHLEAL ETA+I P +NQI LAPGC Q+
    50  Sbjct: 119  TWKEANAQAWQYMEEAVEAGLIKSIGVSNFMVHHLEALQETAKITPAINQIRLAPGCYQK 178

    Query: 200  DLVRFCGNDILLEAYSPFGTGAIFENESIKAIKAEKYGKSVQVALRWSLDNGFLPLPKS 259
              ++V +CK N+ILLEA+SP G G IF+NE+++ +A KY K+VAQVAL WSL GF+PLPKS
    55  Sbjct: 179  EVVDYCKANEILLEAWSPLGQGEIFDNETMQQLANKYDKTVAQVALAWSLAEGFIPLPKS 238

    Query: 260  ATPKNIEANLDIFDFQLNEDDIATLIQLDSGIKPKDPDNVSF 301
              + I+ N+ IFD L ++D T+ L +PD SF
    Sbjct: 239  VHDERIKENMAIFDVSLTQEDKKTIRYLSGMSAIPNPDTSF 280
  
```

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1458-

Example 1324

A DNA sequence (GBSx1406) was identified in *S.agalactiae* <SEQ ID 4055> which encodes the amino acid sequence <SEQ ID 4056>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0633(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10009> which encodes amino acid sequence <SEQ ID 10010> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12612 GB:Z99108 similar to NAD(P)H-flavin oxidoreductase

[Bacillus subtilis]

Identities = 106/223 (47%), Positives = 150/223 (66%), Gaps = 8/223 (3%)

Query: 29 DIKKQVRRAPDFRMAIRVYN--NNDIPKEDMEYILD TAWLSPSSVGLEGWRFVLVDRQTIA 87
D+K Q+ A++FR A + ++ N + D E+IL+T LSPSS+GLE W+F+V+

Sbjct: 3 DLKTQILDAYNFRHATKEFDPNKKVSDSDFEFILETGRLSPSSLGLEPWKFVVVQNP--- 59

Query: 88 KFRDKLKEVAWGAQYQLDTASHFVLLAE--KGAYYNADSMINSLIRRLGLDPAALESRI 145
+FR+KL+E WGAQ QL TASHFVL+LA K YNAD + L E +

Sbjct: 60 EFREKLREYTWGAQKQLPTASHFVLILARTAKDIKYNADYIKRHLKEVKQMPQDVYEGYL 119

Query: 146 PLYKSFOENDMKI--DSERSLWDWTAKQTYIALGNMNTAAAMIGVDS CPIEGFDYKVNNI 204
+ FQ+ND+ + +S+R+L+DW +KQTYIALGNMNTAA IGVDSCPIEGF Y+ ++ I

Sbjct: 120 SKTEBFQKNDLHLLSDRTLFDWASKQTYIALGNMNTAAQIGVDS CPIEGFYDHIHRI 179

Query: 205 LSKEGLIDDKKEAISCMSVSGYRLREP KHSRARKERQEVITWV 247
L +EGL+++ IS MV+FGYR+R+P+ + R ++V+ WV

Sbjct: 180 LEEEGLENGSFDISVMVAFGYRVRDPR-PKTRSAVEDVVKWV 221

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4057> which encodes the amino acid sequence <SEQ ID 4058>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1705(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 126/222 (56%), Positives = 174/222 (77%), Gaps = 4/222 (1%)

Query: 28 EDIKKQVRRAPDFRMAIRVYNNNDIPKEDMEYILD TAWLSPSSVGLEGWRFVLVDRQTIA 87
+ I Q+++A FR A+RVY I ED+ ILD AWLSPSS+GLEGWRF+VLD + I

Sbjct: 3 QTIHQIQQALHFRTAVRVYKEEKISDEDLALILDAWLSPSSIGLEGWRFVVLNKP I- 61

Query: 88 KFRDKLKEVAWGAQYQLDTASHFVLLAEKGAYYNADSMINSLIRRLGLDPAALESRIPL 147
++++K AWGAQYQL+TASHF+LL+AEK A Y++ ++ NSL+RRG+ + L SR+ L

Sbjct: 62 --KEEIKPFAWGAQYQLE+TASHFILLIAEKHARYDSPAIKNSLLRRGIKEGDLNSRLKL 119

Query: 148 YKSFQENDMKI--DSERSLWDWTAKQTYIALGNMNTAAAMIGVDS CPIEGFDYKVNNILS 206
Y+SEQ+ DM + D+ R+L+DWTAKQTYIALGNMM AA++G+D+CPIEGF Y+KVN+IL+

Sbjct: 120 YESFQKEDMDMADNPRALFDWTAKQTYIALGNMNTAALGIDTCPIEGFHYDKVNHILA 179

-1459-

Query: 207 KEGLIDDKKEAISCMSVSGYRLREFKHSRARKERQEVITWVE 248
 K +ID +KE I+ M+S GYRLR+FKH++ RK ++EVI+ V+
 Sbjct: 180 KHNVIDLEKEGIASMLSLGYRLRDPKHAQVRKPKKEEVISVVK 221

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1325

A DNA sequence (GBSx1407) was identified in *S.galactiae* <SEQ ID 4059> which encodes the amino acid sequence <SEQ ID 4060>. This protein is predicted to be lactoylglutathione lyase (gloA). Analysis of
 10 this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1656(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAC21986 GB:U32717 lactoylglutathione lyase (gloA) [Haemophilus influenzae Rd]
 Identities = 59/131 (45%), Positives = 86/131 (65%), Gaps = 2/131 (1%)

Query: 1 MPFLHTCIRVKDLDAIAFYQEALGFKEVRRNDFPENQFTLVYMALEDDPSY-ELELTYN 59
 M LHT +RV DLD SI FYQ+ LG + +R ++ PE ++TL ++ ED S E+ELTYN
 25 Sbjct: 1 MQILHTMLRVGDLDRSIKFYQDVLGMRLRTSENPEYKYTLAFLGYEDGESAAEIELTYN 60

Query: 60 YDHEAYDLGNGYGHIAVGVDLETTYDAHQKAGYSVTKISG-LPGKPNMFYFIQDPDGYK 118
 + + Y+ G YGHIA+GVDD+ T +A + +G +VT+ +G + G + F++DPDGYK
 30 Sbjct: 61 WGVDKYEHGTAYGHIAIGVDDIYATCEAVRASGNGVTREAGFPVKGSTVIAFVEDPDGYK 120

Query: 119 IEVIRLSQFKA 129
 IE I K+
 Sbjct: 121 IEFIENKSTKS 131

- 35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4061> which encodes the amino acid sequence <SEQ ID 4062>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1382(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 80/125 (64%), Positives = 93/125 (74%), Gaps = 1/125 (0%)

Query: 1 MPFLHTCIRVKDLDAIAFYQEALGFKEVRRNDFPENQFTLVYMALEDDPSY-ELELTYN 60
 M LHTCIRVKDL D S+AFY A FKE R DFP++QFTLVY+ALE + SYELELTYN
 50 Sbjct: 1 MKALHTCIRVKDL D QSVAFYTSAPFPKENYRKDFPDSQFTLVYLALEGE-SYELELTYN 59

Query: 61 DHEAYDLGNGYGHIAVGVDLETTYDAHQKAGYSVTKISGLPGKPNMFYFIQDPDGYKIE 120
 H YDLGNGYGHIA+G + E + H++AG+ VT I L K +YFIQDPDGYKIE
 55 Sbjct: 60 GHG DYDLGNGYGHIALGSEHFEADHKHROAGFPVTDIKELADKSARYYFIQDPDGYKIE 119

Query: 121 VIRLS 125
 VI L+
 Sbjct: 120 VIDLN 124

-1460-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1326

- 5 A DNA sequence (GBSx1408) was identified in *S.agalactiae* <SEQ ID 4063> which encodes the amino acid sequence <SEQ ID 4064>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
10  INTEGRAL    Likelihood = -9.02    Transmembrane  241 - 257 ( 229 - 262)
    INTEGRAL    Likelihood = -4.94    Transmembrane  270 - 286 ( 264 - 287)

----- Final Results -----
    bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
15    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB12688 GB:Z99108 stress response protein [Bacillus subtilis]
Identities = 139/304 (45%), Positives = 200/304 (65%), Gaps = 3/304 (0%)
20  Query: 3    LLSVIVPCYNEQETVSTFLTEIKKVESEMARYTHFEYIFVNDGSTDRITLELLKKAQKQFD 62
    L+S+I+P YNE V +KK E + Y +E F+NDGS D TL+ +K A
    Sbjct: 5    LISIIIPSYNEGYNVKLIHESLKK-EFKNIHYD-YEIFFINDGSVDDTLQIKDLAATCS 62

25  Query: 63   NVHYLSFSRHFGRDAALLAGLEHTTGDFITVMDVDLQDPPTLLPEMYLKLQEGYDIVATR 122
    V Y+SFSR+FGK+AA+LAG EH G+ + VMD DLQ P LL E +EGYD V +
    Sbjct: 63   RVKYISFSRNFGKEAAILAGFEHVQGEAVIVMDADLQHPTVLLKEFIKGYREGYDQVIAQ 122

30  Query: 123  RKDRKGEPLIRSLFAKLFYKLINQVSDTKMVDGARDFRLMTKQVVDLSILELNEVNRFSKG 182
    R +RKG+ +RSL + ++YK IN+ + + DG DFRL+++Q V+++L+L+E NRFSKG
    Sbjct: 123  R-NRKGDSFVRSLLSSMYKFKINKAVEVDLRDGVGDPRLLSRQAVNALLKLSEGNRFSKG 181

35  Query: 183  IFSWIGYDVAYISYENRERIAGKTSWSFFNLLKYSLDGFINFSEIPLAIATWIGTLSSVL 242
    +F WIG+D + YEN ER G + WSF +L Y +DG ++F+ PL + + G +L
    Sbjct: 182  LFCWIGFDQKIVFYENVERKNKTSKWSFSSLFNYGMDGVVSNHKEPLRLCFYTGIFILL 241

40  Query: 243  SLLAIIFIIIRKLLFGDPVSGWASTVTIVLFMGGIQLLSLGIIGKYISKIFLETKKRPVY 302
    S++ II ++ L G V G+ + ++ VLF+GG+QLLSLGIIG+YI +I+ ETKKRP Y
    Sbjct: 242  SIIYIIATFVKILTNGISVPGYFTIISAVLFLGGVQLLSLGIIGFYIGRIYYETKKRPHY 301

    Query: 303  IVKE 306
    ++KE
    Sbjct: 302  LIKE 305

```

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4065> which encodes the amino acid sequence <SEQ ID 4066>. Analysis of this protein sequence reveals the following:

```

Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
50  INTEGRAL    Likelihood = -9.55    Transmembrane  256 - 272 ( 251 - 282)
    INTEGRAL    Likelihood = -5.31    Transmembrane  290 - 306 ( 284 - 307)

----- Final Results -----
    bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
55    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9113> which encodes the amino acid sequence <SEQ ID 9114>. Analysis of this protein sequence reveals the following:

-1461-

Possible cleavage site: 36
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5 bacterial membrane --- Certainty= 0.482(Affirmative) < succ>
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 207/307 (67%), Positives = 258/307 (83%)

Query: 1 MALLSVIVPCYNEQETVSTFLTEIKKVESEMARYTHFEYIFVNDGSTDRTELLKKAQK 60
 M LLS+IVPC+NE+ + + E+ ++E+ M FEYIF++DGS D TL +L++ A +
 15 Sbjct: 21 MTLLSIIVPCFNEEANILPYFEEMHQLETSMTNQLAFEYIFIDDGSKDNTLGILRELAAR 80

Query: 61 FDNVHYLSFSRHFQKDAALLAGLEHTTGDFITVMDVDLQDPPTLLPEMYLKLQEGYDIVA 120
 F NVHYLSFSRHFQK+A LLAGL+ G++ITVMDVDLQDPP LLP MY KL+EGYDIV
 20 Sbjct: 81 FPNVHYLSFSRHFQKEAGLLAGLKEAKGNYITVMDVDLQDPPELLPIMYAKLKEGYDIVG 140

Query: 121 TRRKDRKGEPLIRSLFAKLFYKLINQVSDTKMVDGARDFRLMTKQVVDSEILELNEVNRFS 180
 TRR++R+GEPLIRS+ + LFY LI +SDT+MV+G RD+RLMT+QVVDSEILEL EVNRFS
 25 Sbjct: 141 TRRQNRQGEPLIRSMCSNLFYGLIKHLSDEMVGVRDYRLMTRQVVDSEILELGEVNRFS 200

Query: 181 KGIFSWIGVDVAYISYENRRIAGKTSWSFFNLKYSLDGFINFSEIPLAIATWIGTLSS 240
 KGIFSW+GY + Y+S+EN++R GK+ W F+ LL+YSLDGFINFSE+PL IATW GT S
 30 Sbjct: 201 KGIFSWVGVRITYLSFENQKRYGKSRWHFWELLRYSLDGFINFSEMPLEIATWIGTFSF 260

Query: 241 VLSLLAIIFIIIRKLLFGDPVSGWASTVTIVLFMGGIQLLSLGIIGKYISKIFLETKKRP 300
 ++S+ AI+FIIIRK+LFGDPVSGWASTV+I+LFMGGIQL +GIIGKYISKIFLETKKRP
 35 Sbjct: 261 LISIFAILFIIIRKILFGDPVSGWASTVSIILFMGGIQLFCMGIIGKYISKIFLETKKRP 320

Query: 301 VYIVKEE 307
 +YI+KE+
 40 Sbjct: 321 LYIIKEK 327

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1327

40 A DNA sequence (GBSx1409) was identified in *S.agalactiae* <SEQ ID 4067> which encodes the amino acid sequence <SEQ ID 4068>. This protein is predicted to be d-serine/d-alanine/glycine transporter (cycA). Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -2.44 Transmembrane 50 - 66 (50 - 66)
 INTEGRAL Likelihood = -1.49 Transmembrane 27 - 43 (27 - 43)

----- Final Results -----

50 bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA83253 GB:Z31377 potential amino acid permease
 [Lactobacillus delbrueckii]

55 Identities = 34/55 (61%), Positives = 44/55 (79%)

Query: 7 DHTQKSENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSISLTGPSIVLVYAITG 61
 D + ++ +G +R L NRHVQ+IAI GTIGTGLFLGAG +IS TGPS++ +YAI G
 60 Sbjct: 5 DRSIENTDGTIRSLSNRHVQMIAGGTIGTGLFLGAGTTISATGPSVIFYAIMG 59

-1462-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4069> which encodes the amino acid sequence <SEQ ID 4070>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

5	INTEGRAL	Likelihood = -11.15	Transmembrane	170 - 186 (161 - 190)
	INTEGRAL	Likelihood = -8.44	Transmembrane	256 - 272 (252 - 274)
	INTEGRAL	Likelihood = -8.33	Transmembrane	352 - 368 (347 - 375)
	INTEGRAL	Likelihood = -7.54	Transmembrane	139 - 155 (133 - 160)
	INTEGRAL	Likelihood = -5.73	Transmembrane	420 - 436 (417 - 440)
10	INTEGRAL	Likelihood = -3.88	Transmembrane	56 - 72 (54 - 75)
	INTEGRAL	Likelihood = -3.40	Transmembrane	283 - 299 (282 - 300)
	INTEGRAL	Likelihood = -3.29	Transmembrane	440 - 456 (439 - 458)
	INTEGRAL	Likelihood = -1.49	Transmembrane	31 - 47 (31 - 47)
15	INTEGRAL	Likelihood = -1.33	Transmembrane	109 - 125 (109 - 127)

----- Final Results -----
 bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14651 GB:Z99117 amino acid permease [Bacillus subtilis]
 Identities = 210/454 (46%), Positives = 296/454 (64%), Gaps = 11/454 (2%)

25 Query: 12 DNNELENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSIALTGPSIIFVYMITGAFMFMM 71
 DN + + RGL+NRH+QL+AI G IGTGLFLG+G+SI GPSI+F Y+ITG F F +
 Sbjct: 8 DNFGQQQKLSRGLKNRHIQLMAIGGAIGTGLFLGSGKSIHFAGPSILFAYLITGVFCFFI 67

30 Query: 72 MRAIGEMLYYDPDQHTFINFISKYIGPGWGYFSGLSYWISLIFIGMAEITAVGAYVQFWF 131
 +R++GE+L + H+F++F+ Y+G + +G +YW I + MA++TAVG Y Q+W
 Sbjct: 68 IRSLGELLLSNAGYHSFVDFVRDYLGNMAAFITGWTYWFCWISLAMADLTAVGIYTQYWL 127

35 Query: 132 PSWPAWLILQVLVLLSSINLIARVVFGETEFWFAMIKILAILALIATAIFMVLTFETH 191
 P P WL L+ L++L +NL V++FGE EFWFA+IK++AILALI T I ++ GF
 Sbjct: 128 PDVQPWLPGLLALILLIMNLATVKLFGELEFWFALIKVIAIALIVTGILLIAKGFSA 187

40 Query: 192 TGHASLSNIFDHFSMFNGKLFKFFMAFQMVFFAYQAIEFVGITTTSETANPRKVLPAIQE 251
 +G ASL+N++ H MFPNG F ++FQMV FA+ IE VG+T ET NP+KV+PKAI +
 Sbjct: 188 SGPASLNNLWSHGGMFPNGWGHGFIISFQMVVFAFVGIELVGLTAGETENPQKVIPKAINQ 247

45 Query: 252 IPTRIVIFYVGALVSIMAIVPWHQLPVDESFFVMVFKLIGIKWAAALINFVVLTSASAL 311
 IP RI++FYVGAL IM I PW+ L +ESFFV VF +GI AA+LINFVVLTSASAA
 Sbjct: 248 IPVRIILFYVGALFVIMCIYPWNVLPNPFVQVFSAGVIVVAALINFVVLTSASAAA 307

50 Query: 312 NSTLYSTGRHLYQIANE--TPNALTNRLKINTLSROGVPSRAIIASAVVVGISALINILP 369
 NS L+ST R +Y +A + P L L+ VPS A+ S++ + I +N L
 Sbjct: 308 NSALFSTSRMVYSIAKDHAPGLL-----KKLTSNNVPSNALFFSSIALIGVSLNYLM 361

55 Query: 370 GVADAFSLITASSSGVYIAIYALTMIAHWKYRQSK--DFMADGYLMPKYKVTPLTLAFL 427
 F+LIT+ S+ +I I+ +T+I H KYR+++ + A+ + MP Y ++ LTLAF
 Sbjct: 362 -PEQVFTLITSVSTICFIFIWGITVICHLYKRYKTRQHEAKANKFKMPFYPLSNYLTFLAFL 420

Query: 428 AFVFISLFLQESTYIGAIGATIWIIFGIYSNVK 461
 AF+ + L L T I +W ++ I V+
 Sbjct: 421 AFILVILALANDTRIALFVTPVWFVLLIILYKVQ 454

An alignment of the GAS and GBS proteins is shown below.

Identities = 48/62 (77%), Positives = 51/62 (81%)

60 Query: 1 MSKNNNDHTQKSENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSISLTGPSIVLVYAITGA 62
 MS + ENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSI+LTGPSI+ VY ITGA
 Sbjct: 5 MSIKEQTDNNELENGMVRGLENRHVQLIAIAGTIGTGLFLGAGRSIALTGPSIIFVYMITGA 66

-1463-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1328

A DNA sequence (GBSx1411) was identified in *S.agalactiae* <SEQ ID 4071> which encodes the amino acid sequence <SEQ ID 4072>. This protein is predicted to be alkylphosphonate uptake protein (phnA). Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0965(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77069 GB:AE000483 orf, hypothetical protein [Escherichia coli K12]
Identities = 79/110 (71%), Positives = 91/110 (81%), Gaps = 1/110 (0%)

Query: 1 MSLPNCPCNCKSEYVYEDGILLVCPECAEYWNPEE-IEEEVGLIVLDSNGTRLSDGDTVTIV 59
MSLP+CPKCNSY YED + +CPECAEYWN E +E LIV D+NG L+DGD+VT+
Sbjct: 1 MSLPHCPKCNSYTYEDNGMYICPECAEYWNDAEPAQESDELIVKDANGNLLADGDSVTI 60

Query: 60 IKDLKVKGAPKDIKQGTRVKNIRLVGDHNDCKIDGFGAMKLKSEFVKK 109
IKDLKVKG+ +K GT+VKNIRLV+GDHNDCKIDGFG MKLKSEFVKK
Sbjct: 61 IKDLKVKGSSSMLKIGTKVKNIRLVGDHNDCKIDGFGPMKLKSEFVKK 110

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4073> which encodes the amino acid sequence <SEQ ID 4074>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3428(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 73/85 (85%), Positives = 79/85 (92%), Gaps = 1/85 (1%)

Query: 26 CAYEWNPEEIEEEVGLIVLDSNGTRLSDGDTVTIVIKDLKVKGAPKDIKQGTRVKNIRLV 84
CA+EW P EE EE GL+VLDSNG RLSDGDT+TV+KDLKVKGAPKDIKQGTRVKNIRLV
Sbjct: 2 CAFEWTPGEEATEEBEGLVVLDSNGVRLSDGDTTITVVKDLKVKGAPKDLKQGTRVKNIRLV 61

Query: 85 DGDHNDCKIDGFGAMKLKSEFVKK 109
+GDHNDCKIDGFGAMKLKSEFVKK
Sbjct: 62 EGDHNDCKIDGFGAMKLKSEFVKK 86

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1329

A DNA sequence (GBSx1412) was identified in *S.agalactiae* <SEQ ID 4075> which encodes the amino acid sequence <SEQ ID 4076>. Analysis of this protein sequence reveals the following:

Possible site: 22

-1464-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3665(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 500.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1330

A DNA sequence (GBSx1414) was identified in *S.agalactiae* <SEQ ID 4077> which encodes the amino acid sequence <SEQ ID 4078>. Analysis of this protein sequence reveals the following:

15 Possible site: 13

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -0.11 Transmembrane 558 - 574 (558 - 574)

----- Final Results -----

20 bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:CAB11971 GB:Z99105 L-glutamine-D-fructose-6-phosphate
 amidotransferase [Bacillus subtilis]
 Identities = 355/604 (58%), Positives = 445/604 (72%), Gaps = 4/604 (0%)

30 Query: 1 MCGIVGVVGNNTATDILIQGLEKLEYRGYDSAGIFVVGDNKSQLVKS VGR IAEIQAKVGD 60
 MCGIVG +G +A +IL++GLEKLEYRGYDSAGI V + + K GRIA+++ V
 Sbjct: 1 MCGIVGYIGQLDAKEILLKGLEKLEYRGYDSAGIAVANEQGIHVFKREKGR IADLREV VDA 60

35 Query: 61 SVSGTGTIGHTRWATHGKPTTEGNAHPHTSGSGRFVLVHNGVIENYLQIKETYLTKHNLKG 120
 +V GIGHTRWATHG+P+ NAHPH S GRF LVHNGVIENY+Q+K+ YL LK
 Sbjct: 61 NVEAKAGIGHTRWATHGEP SYLNAHPQSALGRFTLVHNGVIENYVQLKQBYLQDVELKS 120

40 Query: 121 ETDTEIAIHLVEHFVEEDNLSVLEAFKKALHIEG SYAFALIDSQDADTIYVAKNKSPIL 180
 +TDTE+ + ++E FV L EAF+K L +++GSYA AL D+ + +TI+VAKNKSPIL
 Sbjct: 121 DTDTEVVVQVIEQFVN-GGLETFEEAPRKTLTLKGSYAIALFDNDNRETIFVAKNKSPIL 179

45 Query: 181 IGLGNGYNMVCSDAMAMIRETSEYMEIHDKELVIVKKDSVEVQDYDGNVIERG SYTAE LD 240
 +GLG+ +N+V SDAMAM++ T+EY+E+ DKE+VIV D V +++ DG+VI R SY AE LD
 Sbjct: 180 VGLGDTFNVVASDAMAMLQVTNEYVELMDKEMVIVTDDQVVIKNLDG D VITRAS YIAE LD 239

50 Query: 241 LSDIGKGTYPFYMLKEIDEQPTVMRKLISTYANESGDMNVDSDI IKS VQEADRL YILAAG 300
 SDI KGTYP YMLKE DEQP VMRK+I TY +E+G ++V DI +V EADR+YI+ G
 Sbjct: 240 ASDIEKGTYPHYMLKETDEQPVMRKIIQTYQDENGKLSVPGDIAAAVAEADRIYIIGCG 299

55 Query: 301 TSYHAGFAAKTMIKLTITPVELGVSSEWGYNMP LLSKKPMPFILLSQSGETADSRQVLVK 360
 TSYHAG K IE + PVE+ V+SE+ YNMP LLSKKP+FI LSQSGETADSR VLV+
 Sbjct: 300 TSYHAGLVGQYIEMWANVPVEVHVASEFSYNMP LLSKKPLFIFLSQSGETADSR AVL VQ 359

60 Query: 361 ANEMGIPSLTITNVP GSTLSREATY TMLIHAGPEIAVASTKAYTAQVATLAF LAKAVGEA 420
 +G +LTITNVP GSTLSREA YT+L+HAGPEIAVASTKAYTAQ+A LA LA +
 Sbjct: 360 VKALGHKALTITNVP GSTLSREADYTLLHAGPEIAVASTKAYTAQIAVLAVLASVAADK 419

Query: 421 NGKAEAKDFDLVHEL SIVAQSIEATLSEKDVISEKVEQLLISTRNAFYIGRGNDYVVTME 480
 NG FDLV EE I A ++EA +KD + + L +RNAF+IGRG DY+V +E
 Sbjct: 420 NGINIG--FDLVKELGIAANAMEALCDQKDEMEMIAREYLT VSRNAFFIGRGLDYFVCVE 477

-1465-

Query: 481 AALKLKEISYIQTEGFAAGELKHGTISLIEDNTPVIALISADSTIAAHTRGNIQEVVSRG 540
 ALKLKEISYIQ EGFA GELKHGTI+LIE TPV AL + + + RGN++EV +RG
 Sbjct: 478 GALKLKEISYIQAEFGAGGELKHGTIALIEQGTVPFALATQEH-VNLSIRGNVKEVAARG 536

5 Query: 541 ANALIIVEEGLEREGDDIIVNKVHPFLSAISMVIPTQLIAYYASLQRLDVKPRNLAKA 600
 AN II +GL+ .D ++ +V+P L+ + V+P QLIAYYA+L RG DVDKPRNLAK+
 Sbjct: 537 ANTCTIISLKGLDDADDRFVLPEVNEPALPLVSVVPLQLIAYYAALHRCDDVDKPRNLAKS 596

10 Query: 601 VTVE 604
 VTVE
 Sbjct: 597 VTVE 600

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4079> which encodes the amino acid sequence <SEQ ID 4080>. Analysis of this protein sequence reveals the following:

15 Possible site: 39
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.06 Transmembrane 558 - 574 (558 - 574)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:CAB11971 GB:Z99105 L-glutamine-D-fructose-6-phosphate
 amidotransferase [Bacillus subtilis]
 Identities = 353/604 (58%), Positives = 445/604 (73%), Gaps = 4/604 (0%)
 30 Query: 1 MCGIVGVGNRNATDILMQGLEKLEYRGYDSAGIFVANANQTNLIKSVGRIADLRKIGI 60
 MCGIVG +G +A +IL++GLEKLEYRGYDSAGI VAN ++ K GRIADLR +
 Sbjct: 1 MCGIVGYIGQLDAKEIILKGLEKLEYRGYDSAGIAVANEQGIHVFKKGRIADLRVVD 60
 Query: 61 DVAGSTGIGHTRWATHGQSTEDNAHPHTSQTGRFVLVHNGVNIENYLHIKTEFLAGHDFKG 120
 +V GIGHTRWATHG+ + NAHPH S GRF LVHNGVNIENY+ +K E+L + K
 35 Sbjct: 61 NVEAKAGIGHTRWATHGEPYSLNAHPHQALGRFTLVHNGVNIENYVQLKQEYLQDVELKS 120
 Query: 121 QTDTEIAVHLIGKFVEEDKLSVLEAFKKSLSIIEGSYAFALMDSQATDTIYVAKNKSPLL 180
 TDTE+ V +I +FV L EAF+K+L++++GSYA AL D+ +TI+VAKNKSPLL
 40 Sbjct: 121 DTDTEVVVQVIEQFVNGG-LETEBAFRKTLTLKGSYAIALFDNDNRETIFVAKNKSPLL 179
 Query: 181 IGLGEGYNMVCSAMAMIRETSEFMEIHKELVILTCKDKVTVDYDGKELIRDSYTAELD 240
 +GLG+ +N+V SDAMAM++ T+E++E+ DKE+VI+T D+V + + DG + R SY AELD
 Sbjct: 180 VGLGDTFNVVASDAMAMLQVTNEYVELMDKEMVIVTDDQVVIKNLDGDTVITRASIAELD 239
 45 Query: 241 LSDIGKGTYPFYMLKEIDEQPTVMRQLISTYADETGNVQVDPAIITSIQEADRLYILAAG 300
 SDI KGTYP YMLKE DEQP VMR++I TY DE G + V I ++ EADR+YI+ G
 Sbjct: 240 ASDIEKGTYPHYMLKETDEQPVVMRKIIQTYQDENGKLSVPGDIAAAVAEADRIYIICG 299
 Query: 301 TSYHAGFATKNMLEQLTDTPELVGVAEWSGYHMPILLSKKPMFILLSQSGETADSRQVLVK 360
 TSYHAG K +E + PVE+ VASE+ Y+MPILLSKKP+FI LSQSGETADSR VLV+
 50 Sbjct: 300 TSYHAGLVGKQYIEMWANVPVEVHVASEFSYNMPLLSKKPLFIPLSQSGETADSRVILVQ 359
 Query: 361 ANAMGIPSLTNTVPGSTLSREATYTMILHAGPEIAVASTKAYTAQIAALAFKAVGEA 420
 A+G +LT+TNVPGSTLSREA YT+L+HAGPEIAVASTKAYTAQIA LA LA +
 55 Sbjct: 360 VKALGHKALTITNVPGSTLSREADYTLLHAGPEIAVASTKAYTAQIAVLAVLASVAADK 419
 Query: 421 NGKQEAALDFNLVHLSLVAQSIEATLSEKDLVAEKVQALLATRNFAFYIGRGNDYVAME 480
 NG + F+LV EL + A ++EA +KD + + L +RNAF+IGRG DY+V +E
 60 Sbjct: 420 NGIN--IGFDLVKELGIAANAMEALCDQKDEMELAREYLTVSRNAFFIGRGDLYFVCVE 477
 Query: 481 AALKLKEISYIQCEGFAAGELKHGTISLIEEDTPVIALISSQLVASHTRGNIQEVAAARG 540
 ALKLKEISYIQ EGFA GELKHGTI+LIE+ TPV AL + + S RGN++EVAARG
 Sbjct: 478 GALKLKEISYIQAEFGAGGELKHGTIALIEQGTVPFALATQEHVNLS-IRGNVKEVAARG 536
 65 Query: 541 AHVLTVEEGLDREGDDIIVNKVHPFLAPIAMVIPTQLIAYYASLQRLDVKPRNLAKA 600

-1466-

A+ + +GLD D ++ +V+P LAP+ V+P QLIAYYA+L RG DVDKPRNLAK+
 Sbjct: 537 ANTCTISLKGLEDADRFLVPEVNPALAPLVSVPLQLIAYYAALHRGCDVDKPRNLAKS 596

Query: 601 VTVE 604
 VTVE
 Sbjct: 597 VTVE 600

An alignment of the GAS and GBS proteins is shown below.

Identities = 500/604 (82%), Positives = 552/604 (90%)

Query: 1 MCGIVGVVGNNTATDILIQGLEKLEYRGYDSAGIFVVGDNKSQLVKSVGRIAETQAKVGD 60
 MCGIVGVVGN NATDIL+QGLEKLEYRGYDSAGIFV N++ L+KSVGRIA+++AK+G
 Sbjct: 1 MCGIVGVVGNRNATDILMQGLEKLEYRGYDSAGIFVANANQTNLIKSVGRADLRAKIGI 60

Query: 61 SVSGTTGIGHTRWATHGKPTTEGNAHPHTSGSGRFVLVHNGVIENYLQIKETYLTKHNLKG 120
 V+G+TGIGHTRWATHG+ TE NAHPHTS +GRFVLVHNGVIENYL IK +L H+ KG
 Sbjct: 61 DVAGSTGIGHTRWATHGQSTEDNAHPHTSQTGRFVLVHNGVIENYLHIKTEFLAGHDFKG 120

Query: 121 ETDTEIAIHLVEHFVEEDNLSVLEAFKKALHIEGSAFALIDSQADTTIYVAKNKSPLL 180
 +TDTEIA+HL+ FVEED LSVLEAFKK+L IIEGSAFAL+DSQ DTIYVAKNKSPLL
 Sbjct: 121 QTDTEIAVHLIGKFVEEDKLSVLEAFKKSLIIEGSAFALMDSQATDTIYVAKNKSPLL 180

Query: 181 IGLGNGYNMVCSAMAMIRETSEYMEIHDKELVIVKKDSVEVQDYDGNVIERGSAETALD 240
 IGLG GYNMVCSAMAMIRETSE+MEIHDKELVI+ KD V V DYDG + R SYTALD
 Sbjct: 181 IGLGEGYNMVCSAMAMIRETSEFMEIHDKELVILTCKDKVTVDYDGKELIRDSYTALD 240

Query: 241 LSDIGKGTYPFMYLKEIDEQPTVMRKLITYANESGDMNVDSDIKSVQEADRLYILAAG 300
 LSDIGKGTYPFMYLKEIDEQPTVMR+LISTYA+E+G++ VD II S+QEADRLYILAAG
 Sbjct: 241 LSDIGKGTYPFMYLKEIDEQPTVMRQLITYADETGNVQVDPAAITSIQEADRLYILAAG 300

Query: 301 TSYHAGFAAKTMIEKLTDTPELVGSSEWGYNMPLLSKKPMFILLSQSGETADSRQVLVK 360
 TSYHAGFA K M+E+LTDTPVELGV+SEWGY+MPLLSSKKPMFILLSQSGETADSRQVLVK
 Sbjct: 301 TSYHAGFATKNMLEQLTDTPELVGVASEWGYHMPLLSSKKPMFILLSQSGETADSRQVLVK 360

Query: 361 ANEMGIPSLTITNVPGSTLSREATYTMLIHAGPEIAVASTKAYTAQVATLAFLAKAVGEA 420
 AN MGIPSLT+TNVPGSTLSREATYTMLIHAGPEIAVASTKAYTAQ+A LAFLAKAVGEA
 Sbjct: 361 ANAMGIPSLTITNVPGSTLSREATYTMLIHAGPEIAVASTKAYTAQIAALAFLAKAVGEA 420

Query: 421 NGKARAKDFDLVHELSSIVAQSIEATLSEKDVISEKVEQLLISTRNAFYIGRGNDYYVTME 480
 NGK EA DF+LVHELSS+VAQSIEATLSEK+++EKV+ LL +TRNAFYIGRGNDYYV ME
 Sbjct: 421 NGKQEAALDFNLVHELSSIVAQSIEATLSEKDLVAEKVQALLATRNAFYIGRGNDYYVAME 480

Query: 481 AALKLKEISYIQTEGFAAGELKHGTISLIEDNTPVIALISADSTIAAHTRGNIEVVSRG 540
 AALKLKEISYIQ EGFAAGELKHGTISLIE++TPVIALIS+ +A+HTRGNIEV +RG
 Sbjct: 481 AALKLKEISYIQCEGFAAGELKHGTISLIEEDTPVIALISSQLVASHTRGNIEVAARG 540

Query: 541 ANALIIVEEGLEREGLDDIIVNKVHPFLSAISMVPTQLIAYYASLQRLDVKPRNLAKA 600
 A+ L +VEEGL+REGDDIIVNKVHPFL+ I+MVIPTQLIAYYASLQRLDVKPRNLAKA
 Sbjct: 541 AHVLTVVEEGLDREGDDIIVNKVHPFLAPIAMVPTQLIAYYASLQRLDVKPRNLAKA 600

Query: 601 VTVE 604
 VTVE
 Sbjct: 601 VTVE 604

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1331

A DNA sequence (GBSx1415) was identified in *S.agalactiae* <SEQ ID 4081> which encodes the amino acid sequence <SEQ ID 4082>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have a cleavable N-term signal seq.

-1467-

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9797> which encodes amino acid sequence <SEQ ID 9798> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44435 GB:U65000 type-I signal peptidase SpsB [Staphylococcus aureus]
 Identities = 62/185 (33%), Positives = 97/185 (51%), Gaps = 12/185 (6%)
 Query: 10 VKRDFIRNIILALIAVLILILLRYFVFATFKVHKDATNSYFSNGDVVVVN---RNRTPK 65
 +K++ + II +A +IL ++ F+ + + ++ + +G+ V VN + +
 Sbjct: 1 MKKELLEWIISIYAVFVILFIVGKFIVTPYTIKGESMDPTLKDGERVAVNIIGYKTGGLE 60
 Query: 66 YKDFIVYKVGKIF-YISRVIGEPNQKVRVMDLILYLNDFKDEPYIEKMKNAYSEKKDGQ 124
 + +V+ K Y+ RVIG P KV +D LY+N +DEPY+ N + K G
 Sbjct: 61 KGNVVVFHANKNDYVVRVIGVPGDKVEYKNDTLYVNGKKQDEPYL---NYNLKHKQGD 116
 Query: 125 MPFTSDFSVELT--TRNKESRVPKGSYLVLDNRQNKNSRKFGLIKEKDIRGVITFKVY 182
 T F V+ L K + +PKG YLVL DNR+ DSR FGLI E I G ++F+ +
 Sbjct: 117 Y-ITGTFQVKDLPNANPKSNVIPKGKYLVLGDNREVS KDSRAFGLEDQIVGKVSFRFW 175
 Query: 183 PLSEF 187
 P SEF
 Sbjct: 176 PFSEF 180

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4083> which encodes the amino acid sequence <SEQ ID 4084>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood =-14.22 Transmembrane 10 - 26 (4 - 34)

----- Final Results -----

bacterial membrane --- Certainty=0.6689(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 99/185 (53%), Positives = 130/185 (69%)
 Query: 9 MVKRDIFIRNIILALIAVLILILLRYFVFATFKVHKDATNSYFSNGDVVVVNRNRTPKYKD 68
 MVKRDIFIRNI+L LI ++ ILLR FVF+TFKV + N+Y +GD+V + +N PKYKD
 Sbjct: 1 MVKRDIFIRNILLIIVIGAILLRIFVSTFKVSPETANTYKSGDLVTIKNIQPKYKD 60
 Query: 69 FIVYKVGKIFYISRVIGEPNQKVRVMDLILYLNDFKDEPYIEKMKNAYSEKKDGQMPFT 128
 F+VY+VGK Y+SRVI V MDDI YLN++ + + Y+EKMK Y +T
 Sbjct: 61 FVVYRVGKKDYVSRVIAVEGDSVTYMDIFYLNNMVESQAYLEKMKAHYLNHAPFGTLYT 120
 Query: 129 SDFSVELTTRNKESRVPKGSYLVLDNRQNKNSRKFGLIKEKDIRGVITFKVYPLSEFG 188
 DF+V T+T +K +VPKG YL+LNDNR+N NDSR+FGLI I+G++TF+V PLS+FG
 Sbjct: 121 DDFTVATITADKYQKVPKGKYLILLNDNRKNTNDSRRFGLINASQIKGLVTFRVLEPLSDFG 180
 Query: 189 FTASE 193
 F E
 Sbjct: 181 FVEVE 185

A related GBS gene <SEQ ID 8789> and protein <SEQ ID 8790> were also identified. Analysis of this protein sequence reveals the following:

-1468-

Lipop: Possible site: -1 Crend: 10

McG: Discrim Score: 10.13

GvH: Signal Score (-7.5): 0.45

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 0 value: 3.82 threshold: 0.0

PERIPHERAL Likelihood = 3.82 69

modified ALOM score: -1.26

*** Reasoning Step: 3

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

36.0/59.9% over 165aa

Bacillus caldolyticus

EGAD|24914| signal peptidase i Insert characterized

ORF00169(364 - 867 of 1179)

EGAD|24914|25718(15 - 180 of 182) signal peptidase i {Bacillus caldolyticus}

%Match = 11.9

%Identity = 35.9 %Similarity = 59.9

Matches = 60 Mismatches = 61 Conservative Sub.s = 40

```

312      342      372      402      432      462      483      510
L*KHDIMEKRLGVVMVKRDFIRNIILALIAVLILILLRYFVFATFKVHKDATNSYFSNGDVVVVNR---NRTPKYK-DFI
      |  ::      :|| :| |||: :| : : :|:::|: : | : | |
      VTKQKEKRGRWPWFVAVCVVATLRLFVFSNYVVEGKSMMPTLESGNLLIVNKLSDIGPIRRFDII
              10      20      30      40      50      60

537      567      597      627      657      687      717      747
VYKVGKIF-YISRVIGEPNQKVRVMDILYLNDVFKDEPYIEKMKNAYSEKKDQMPFTSDFSVETLTRNKESRVPKGSY
|: | | :| ||| | : : :|||:| |||: | : ||:: | ||::| :| :||| |
VFHANKKEDYVKRVIGLPGDRIAYKNDILYVNGKKVDEPYLRPYKQ---KLLDGR--TGDFTLEEV--GKTRVPPGCI
      80      90      100      110      120      130      140

777      807      837      867      897      927      957      987
LVLNDNRQNKNDNRKFLIKEKDIRGVITFKVYPLSEFGFTASE**KNGII*YHSFVIKWLNRNIF*DR*NF**RXN*
:| | ||| : ||| ||::| | | :| : :| :| |
FVLGDNRLSSWDSRHFVFKINQIVGKVDFRYWPFKQFAFQF
      150      160      170      180

```

SEQ ID 8790 (GBS7) was expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 4; MW 46kDa). It was also expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 4; MW 21kDa). The GBS7-His fusion product was purified (Figure 189, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 262), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1332

A DNA sequence (GBSx1416) was identified in *S. agalactiae* <SEQ ID 4085> which encodes the amino acid sequence <SEQ ID 4086>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

-1469-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1099(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9795> which encodes amino acid sequence <SEQ ID 9796> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF25804 GB:AF172173 pyruvate kinase [Streptococcus thermophilus]
 Identities = 413/500 (82%), Positives = 451/500 (89%)

Query: 1 MNKRKIVATLGPVAVFRGGKIFGESGYWGESLDVEASAEKIAQLIKEGANVFRNFNSHG 60
 MNKRKIVATLGPVAVFRGGKIFGESGYW E LD +ASA+ IAQLI+EGANVFRNFNSHG
 Sbjct: 1 MNKRKIVATLGPVAVFRGGKIFGESGYWSEKLDPAKAKNIAQLIEEGANVFRNFNSHG 60

Query: 61 DHAEQSGARMATVRKABEIIAGQKVGFLDDTKGPEIRTELFEDGADFHSYTTGTLRVATKQ 120
 +HAEQG RM VR AE IAGQKVGFLDDTKGPEIRTELF E A ++Y TG ++R+ATKQ
 Sbjct: 61 NHAEQGERMDVVRMAESIIAGQKVGFLDDTKGPEIRTELFEGDAKEYAYKTGEQIRIATKQ 120

Query: 121 GIKSTPEVIALNVAGGLDIFDDVEVGKQILVDDGKGLTVFAKDKDTREFEVVENDGLI 180
 G+KST +VIALNVAG LDIFDDVEVGKQ+LVDDGKGL V KD + REF V VENDG+I
 Sbjct: 121 GLKSTRDVIALNVAGALDIFDDVEVGKQVLVDDGKGLRVVDKDAEKREFIVEVENDGII 180

Query: 181 GKQKGVNIPYTKIPFPALAERDNADIRFGLEQGINFIAISFVRTAKDVNEVRAICEETGN 240
 KQKGVNIPYTKIPFPALAERDNADIRFGLEQGINFIAISFVRTAKDV EVRAICEETGN
 Sbjct: 181 AKQKGVNIPYTKIPFPALAERDNADIRFGLEQGINFIAISFVRTAKDVQEVRAICEETGN 240

Query: 241 GHVKLFAKIENQQGIDNIDEIIEAADGIMTARGDMGIEVPFEMVPVYQKMIITKVNAAGK 300
 GHVKL AKIENQQGIDNIDEIIEAADGIMTARGDMGIEVPFEMVPVYQKMIITKVNAAGK
 Sbjct: 241 GHVKLLAKIENQQGIDNIDEIIEAADGIMTARGDMGIEVPFEMVPVYQKMIITKVNAAGK 300

Query: 301 AVITATNMLETMTDKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID 360
 V+TATNMLETMT+KPRATRSEVSDVFNAVIDGTDATMLSGESANG YPVESVRTMATI
 Sbjct: 301 LVVTATNMLETMTDKPRATRSEVSDVFNAVIDGTDATMLSGESANGPYPVESVRTMATIH 360

Query: 361 KNAQTLLNEYGRILDSSAFFRNKNTDVIASAVKDATSMIDIKLVVTTITETGNTARAISKFR 420
 KNAQTLL EYGRIL+SS F R++ T+V+ASAVKDAT+SM I+L+V +TE+GNTA I +R
 Sbjct: 361 KNAQTLLKEYGRILNSSTFDRSSNTEVVASAVKDATNSMHIQLIVALTESGNTASLIDTYR 420

Query: 421 PDADILAVTFDEKVQSLMINWGVIPVLADKPASTDDMFVAERVALEAGFVESGDNIVI 480
 P+ADI A+TFDE Q+SLM+NWGVIPV+ + P+STDDMFVAERVALE+G VESGDNIVI
 Sbjct: 421 PEADIWATTFDELTQKSLMLNWGVIPVVTETPSSTDDMFVAERVALESLVESGDNIVI 480

Query: 481 VAGVPVGTGGTNTMRVRTVK 500
 VAGVPVG+G TNTMR+RTVK
 Sbjct: 481 VAGVPVGSNTNTMRIRTVK 500

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4087> which encodes the amino acid sequence <SEQ ID 4088>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0915(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 272-274

The protein has homology with the following sequences in the databases:

>GP:AAF25804 GB:AF172173 pyruvate kinase [Streptococcus thermophilus]
 Identities = 404/500 (80%), Positives = 457/500 (90%)

-1470-

5
Query: 1 MNKRVKIVATLGPAVEIRGGKKYGEDGYWAGQLDVEESAKIAELIEAGANVFRFNFSHG 60
Sbjct: 1 MNKRVKIVATLGPAVEIRGGKK+GEDGYW+ +LD + SAK IA+LIE GANVFRFNFSHG 60

10
Query: 61 DHKEQGDRLMATVRLAEELIARQKVGFLLDTKGPEMRTELFADDAKEFSYVTGEKIRVATTQ 120
+H EQG+RM VR+AE IA QKVGFLLDTKGPE+RTELF DAKE++Y TGE+IR+AT Q
Sbjct: 61 NHAEQGERMDVVRMAESIAGQKVGFLLDTKGPEIRTELFEGDAKEYAYKTGEQIRIATKQ 120

15
Query: 121 GIQSTRDVIALNVAGSLDIYDEVEVGHTILIDGKGLKVIDKDIATRQFIVEVENDGII 180
G++STRDVIALNVAG+LDI+D+VEVG +L+DDGKGL+V+DKD R+FIVEVENDGII
Sbjct: 121 GLKSTRDVIALNVAGALDIFDDVEVGKQVLVDDGKGLRVVDKDAEKREFIVEVENDGII 180

20
Query: 181 AKQKGVNIPNTKIPFPALAEERDADIRFGLEQGLNFIAISFVRTAKDVEEVREICRETGN 240
AKQKGVNIP TKIPFPALAEERDADIRFGLEQG+NFAISFVRTAKDV+EVR IC ETGN
Sbjct: 181 AKQKGVNIPYTKIPFPALAEERDADIRFGLEQGLNFIAISFVRTAKDVQEVRAICEETGN 240

25
Query: 241 DHVQLFAKIENQQGIDNLDIEIEAADGIMIARGDMGIEVPPFEMVVPVQKMIITKVNAAGK 300
HV+L AKIENQQGIDN+DEIEAADGIMIARGDMGIEVPPFEMVVPV+QKMIITKVNAAGK
Sbjct: 241 GHVKLLAKIENQQGIDNLDIEIEAADGIMIARGDMGIEVPPFEMVVPVQKMIITKVNAAGK 300

30
Query: 301 AVITATNMLETMTTEKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID 360
V+TATNMLETMTTEKPRATRSEVSDVFNAVIDGTDATMLSGESANG YPVESVRTMATI
Sbjct: 301 IIVTATNMLETMTTEKPRATRSEVSDVFNAVIDGTDATMLSGESANGPYPVESVRTMATIH 360

35
Query: 361 RNAQTLLENEYGRLDSSAFPRNKTDVIASAVKDATHSMDIKLVVTTITETGNTARAISKFR 420
+NAQTLLE NEYGR+SS F R++ T+V+ASAVKDAT+SM I+L+V +TE+GNTA I +R
Sbjct: 361 KNAQTLLENEYGRLDSSAFPRNNTDVIASAVKDATHSMDIKLVVTTITETGNTARAISKFR 420

40
Query: 421 PDADILAVTFDEKVRALMINWGVIPVLAEKPASTDMMFEVAERVAEAGLVQSGDNIVI 480
P+ADI A+TFDE Q++LM+NWGVIPV+ E P+STDMMFEVAERVA+E+GLV+SGDNIVI
Sbjct: 421 PEADILAVTFDELTQKSLMLNWGVIPVVTETPSTDDMMFEVAERVAESGLVSGDNIVI 480

45
Query: 481 VAGVPVGTGGTNTMRVRTVK 500
VAGVPVG+G TNTMR+RTVK
Sbjct: 481 VAGVPVGSNTNTMRVRTVK 500

An alignment of the GAS and GBS proteins is shown below.

40
Identities = 440/500 (88%), Positives = 462/500 (92%)

45
Query: 1 MNKRVKIVATLGPAVEFRGGKKFGESGYWGESLDVEASAELIAQLIKEGANVFRFNFSHG 60
MNKRVKIVATLGPAVE RGGKK+GE GYW LDVE SA+KIA+LI+ GANVFRFNFSHG
Sbjct: 1 MNKRVKIVATLGPAVEIRGGKKYGEDGYWAGQLDVEESAKIAELIEAGANVFRFNFSHG 60

50
Query: 61 DHAEQARMATVRKAEEIAGQKVGFLLDTKGPEIRTELFEDGADFHSYTTGTLRVATKQ 120
DH EQG RMA TVR AEEIA QKVGFLLDTKGPE+RTELF D A SY TG K+RVAT Q
Sbjct: 61 DHKEQGDRLMATVRLAEELIARQKVGFLLDTKGPEMRTELFADDAKEFSYVTGEKIRVATTQ 120

55
Query: 121 GIKSTPEVIALNVAGGLDIFDDVEVGKQILVDDGKGLGLTVFAKDKDTREFEVVENDGLI 180
GI+ST +VIALNVAG LDI+D+VEVG IL+DDGKGL V KD TR+F V VENDG+I
Sbjct: 121 GIQSTRDVIALNVAGSLDIYDEVEVGHTILIDGKGLKVIDKDIATRQFIVEVENDGII 180

60
Query: 181 GKQKGVNIPYTKIPFPALAEERDADIRFGLEQGLNFIAISFVRTAKDVNEVRAICEETGN 240
KQKGVNIP TKIPFPALAEERDADIRFGLEQGLNFIAISFVRTAKDV EVR IC ETGN
Sbjct: 181 AKQKGVNIPNTKIPFPALAEERDADIRFGLEQGLNFIAISFVRTAKDVEEVREICRETGN 240

65
Query: 241 GHVKLFKAKIENQQGIDNLDIEIEAADGIMIARGDMGIEVPPFEMVVPVQKMIITKVNAAGK 300
HV+LFKAKIENQQGIDN+DEIEAADGIMIARGDMGIEVPPFEMVVPV+QKMIITKVNAAGK
Sbjct: 241 DHVQLFAKIENQQGIDNLDIEIEAADGIMIARGDMGIEVPPFEMVVPVQKMIITKVNAAGK 300

70
Query: 301 AVITATNMLETMTDKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID 360
AVITATNMLETMT+KPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID
Sbjct: 301 AVITATNMLETMTTEKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID 360

75
Query: 361 KNAQTLLENEYGRLDSSAFPRNNTDVIASAVKDATHSMDIKLVVTTITETGNTARAISKFR 420
+NAQTLLENEYGRLDSSAFPR NKT DVIASAVKDATHSMDIKLVVTTITETGNTARAISKFR
Sbjct: 361 RNAQTLLENEYGRLDSSAFPRNNTDVIASAVKDATHSMDIKLVVTTITETGNTARAISKFR 420

-1471-

Query: 421 PDADILAVTFDEKVRSLMINWGVIPVLADKPASTDDMFVAERVALEAGFVESGDNIVI 480
 PDADILAVTFDEKVR+LMINWGVIPVLA+KPASTDDMFVAERVA+EAG V+SGDNIVI
 Sbjct: 421 PDADILAVTFDEKVRSLMINWGVIPVLA+KPASTDDMFVAERVAEAGLVQSGDNIVI 480

Query: 481 VAGVPVGTGGTNTMRVRTVK 500
 VAGVPVGTGGTNTMRVRTVK
 Sbjct: 481 VAGVPVGTGGTNTMRVRTVK 500

10 A related GBS gene <SEQ ID 8791> and protein <SEQ ID 8792> were also identified. Analysis of this protein sequence reveals the following:

Belongs to Glycolysis/gluconeogenesis pathway. Proteins belonging to this metabolic pathway have been experimentally detected on the surface of *Streptococci*.

The protein has homology with the following sequences in the databases:

15 >GP|6708108|gb|AAF25804.1|AF172173_2|AF172173 pyruvate kinase
 {*Streptococcus thermophilus*}

Score = 821 bits (2098), Expect = 0.0
 Identities = 412/500 (82%), Positives = 450/500 (89%)

20 Query: 1 MNKRKIVATLGPFAVEFRGGKKFGESGYWGESLDVEASAQKIAQLIKEGANVFRFNFSHG 60
 MNKRKIVATLGPFAVE RGGKKFGE GYW E LD +ASA+ IAQLI+EGANVFRFNFSHG
 Sbjct: 1 MNKRKIVATLGPFAVEIRGGKKFGEDGYWSEKLDPDASAKNIAQLIEGANVFRFNFSHG 60

25 Query: 61 DHAEQGMATVRKAEEIAGQKVGFLLDTKGPEIRTELFEADGFHSYTTGTLRVATKQ 120
 +HAEQG RM VR AE IAGQKVGFLLDTKGPEIRTELE A ++Y TG ++R+ATKQ
 Sbjct: 61 NHAEQGERMDVVRMAESIAGQKVGFLLDTKGPEIRTELEFGDAKEYAYKTGQIRIATKQ 120

30 Query: 121 GIKSTPEVIALNVAGGLDIFDDVEVGKQILVDDGKLGTLVFAKDKDTREFEVVENDGLI 180
 G+KST +VIALNVAG LDIFDDVEVGKQ+LVDDGKLG V KD + REF V VENDG+I
 Sbjct: 121 GLKSTRDVIALNVAGALDIFDDVEVGKQVLVDDGKGLRVVDKDAEKREFIVEVENDGII 180

35 Query: 181 GKQKGVNIPYTKIPFPALAERDNADIRFGLQGLNFIAISFVRTAKDVNEVRAICEETGX 240
 KQKGVNIPYTKIPFPALAERDNADIRFGLQGLNFIAISFVRTAKDV EVRAICEETG
 Sbjct: 181 AKQKGVNIPYTKIPFPALAERDNADIRFGLQGLNFIAISFVRTAKDVQEVRAICEETGN 240

40 Query: 241 GHVKLFAKIENQQGIDNIDEIIEAADGIMIARGDMGIEVPPFEMVPVYQKMIITKVNAAGK 300
 GHVKL AKIENQQGIDNIDEIIEAADGIMIARGDMGIEVPPFEMVPVYQKMIITKVNAAGK
 Sbjct: 241 GHVKLLAKIENQQGIDNIDEIIEAADGIMIARGDMGIEVPPFEMVPVYQKMIITKVNAAGK 300

45 Query: 301 AVITATNMLETMTDKPRATRSEVSDVFNAVIDGTDATMLSGESANGKYPVESVRTMATID 360
 V+TATNMLETMT+KPRATRSEVSDVFNAVIDGTDATMLSGESANG YPVESVRTMATI
 Sbjct: 301 IVVTATNMLETMTDKPRATRSEVSDVFNAVIDGTDATMLSGESANGPYPVESVRTMATIH 360

50 Query: 361 KNAQTLLNEYGRLDSSAFPPNNKTDVIAAVKDATSM I+L+V +TE+GNTA I +R
 KNAQTLL EYGR+SS F R++ T+V+ASAVKDAT+SM I+L+V +TE+GNTA I +R
 Sbjct: 361 KNAQTLLKEYGRLDSSSTFDRSSNTEVVASAVKDATNSMHIQLIVALTESGNTASLIDTYR 420

55 Query: 421 PDADILAVTFDEKVRSLMINWGVIPVLADKPASTDDMFVAERVALEAGFVESGDNIVI 480
 P+ADI A+TFDE Q+SLM+NWGVIPV+ + P+STDDMFVAERVALE+G VESGDNIVI
 Sbjct: 421 PEADIWAITFDELTQKSLMLNWGVIPVVTETPSSTDDMFVAERVALESGLVESGDNIVI 480

Query: 481 VAGVPVGTGGTNTMRVRTVK 500
 VAGVPVGTGGTNTMRVRTVK
 Sbjct: 481 VAGVPVGTGGTNTMRVRTVK 500

SEQ ID 8792 (GBS330) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 5; MW 59kDa).

GBS330-His was purified as shown in Figure 213, lane 6.

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1472-

Example 1333

A DNA sequence (GBSx1417) was identified in *S.agalactiae* <SEQ ID 4089> which encodes the amino acid sequence <SEQ ID 4090>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0632(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF25803 GB:AF172173 phosphofructokinase [Streptococcus thermophilus]
Identities = 270/337 (80%), Positives = 302/337 (89%), Gaps = 1/337 (0%)
15
Query: 1  MKRIAVLTSGGDAPGMNAAIRAVVRKAISEGMEVYGINQGYGMVTGDIFFLDANSVGGDT 60
      MKRIAVLTSGGDAPGMNAA+RAVV KAISEG+EV+GIN+GY GMV GDIF LDA  V +
Sbjct: 1  MKRIAVLTSGGDAPGMNAAVRAVVLKAISEGIEVFGINRGYAGMVEGDIFKLDKRVENI 60

20
Query: 61  INRGGTFILRSARYPEFAELEGQLKGIEQLKKHIGEVVVIGDGSYHGAMRLTEHGFPVAV 120
      ++RGGTFL+SARYPEFA+LEGQLKGIEQLKK+GIEGVVVIGDGSYHGAMRLTEHGFPVAV
Sbjct: 61  LSRGGTFLQSARYPEFAQLEGQLKGIEQLKKYIGEVVVIGDGSYHGAMRLTEHGFPVAV 120

25
Query: 121 GLPGTIDNDIVGTDYTIIGFDTAVATAVENLDRLDTSASHNRTFVVEVMGRNAGDIALWS 180
      GLPGTIDNDIVGTDYTIIGFDTAVATA E LD++DT+ SH RTFVVEVMGRNAGDIALW+
Sbjct: 121 GLPGTIDNDIVGTDYTIIGFDTAVATATEALDKIQDTAFSHGRFTFVVEVMGRNAGDIALWA 180

30
Query: 181 GIAAGADQIIVPEEEFNIDEVSVNVRAGYAAG-KHHQIIVLAEGVMMSGDEFKTMKAAGD 239
      GIA+GADQIIVPEEE++I+EVV  V+ GY +G K H IIVLAEGVM +EFA MK AGD
Sbjct: 181 GIASGADQIIVPEEEDINEVVRKVKEGYESGEKSHHIIVLAEGVMGAEEFAAKMKEAGD 240

35
Query: 240 DSDLRVITNLGHLRGGSPRTARDRVLASRMGAYAVQLLKEGRGGLAVGVHNEEMVESPILG 299
      SDLR TNLGH++RGGSPRTARDRVLAS MGA+AV LLKEG GG+AVG+HNE++VESPILG
Sbjct: 241 TSDLRATNLGHVIRGGSPRTARDRVLASWMAHAVDLLKEGIGGVAVGHIHNEQLVESPILG 300

Query: 300 LAEEGALFSLTDEGKIVVNNPHKADRLRLAALNRDLAN 336
      AEEGALFSLT++GKI+VNNPHKA L A LNR LAN
Sbjct: 301 TAEAGALFSLTEDGKIVVNNPHKARLDFAEINRSLAN 337

```

40 Proteins in the glycolysis/gluconeogenesis pathway have been experimentally detected on the surface of *Streptococci*.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4091> which encodes the amino acid sequence <SEQ ID 4092>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0632(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 274/336 (81%), Positives = 306/336 (90%), Gaps = 1/336 (0%)
55
Query: 1  MKRIAVLTSGGDAPGMNAAIRAVVRKAISEGMEVYGINQGYGMVTGDIFFLDANSVGGDT 60
      MKRIAVLTSGGDAPGMNAAIRAVVRKAISEGMEVYGIN+GY GMV GDIFPL +  VGD
Sbjct: 1  MKRIAVLTSGGDAPGMNAAIRAVVRKAISEGMEVYGINRGYAGMVDGDIFFLGSKVEVGDK 60

Query: 61  INRGGTFILRSARYPEFAELEGQLKGIEQLKKHIGEVVVIGDGSYHGAMRLTEHGFPVAV 120

```

-1473-

I+RGGTFL SARYPEFA+LEGQL GIEQLKKHGIEGVVIGGDSYHGAMRLTEHGFAV
 Sb jct: 61 ISRGGTFLYSARYPEFAQLEGQLAGIEQLKKHGIEGVVIGGDSYHGAMRLTEHGFAV 120
 Query: 121 GLPGTIDNDIVGTDYTIGFDTAVATAVENLDRLRDTSSASHNRTFVVEVMGRNAGDIALWS 180
 5 G+PGTIDNDI GTDYTIGFDTAV TAVE +D+LRDTS+SH RTFVVEVMGRNAGDIALW+
 Sb jct: 121 GIPGTIDNDIAGTDYTIGFDTAVNTAVEAIDKLRTSSSHGRTFVVEVMGRNAGDIALWA 180
 Query: 181 GIAAGADQIIVPEEEFNIDEVVSNNVRAGYA-AGKHHQIIIVLAEGVMSGDEFKATMKAAGD 239
 GIA+GADQIIVPEEEF+I++V S ++ + GK+H IIVLAEGVMSG+ FA+ +K AGD
 10 Sb jct: 181 GIASGADQIIVPEEEFDIEKVASTIQYDFEHKGKNNHIIIVLAEGVMSGEAFAQKLKEAGD 240
 Query: 240 DSDLRVTNLGHLLRGGSPRTARDRVLASRMGAYAVQLLKEGRGGLAVGVHNEEMVESPIIG 299
 SDLRVTNLGH+LRGGSPRTARDRV+AS MG++AV+LLK+G+GGLAVG+HNEE+VESPIIG
 Sb jct: 241 KSDLRVTNLGHILRGGSPRTARDRVIASWMGSHAVELLKDGKGLAVGIHNEELVESPIIG 300
 15 Query: 300 LAEEGALFSLTDEGKIVVNNPHKADLRALNRDLA 335
 AEEGALFSLT+EGKI+VNNPHKA L AALNR L+
 Sb jct: 301 TAEAGALFSLTEGKIVVNNPHKARLDFAALNRSL 336

20 SEQ ID 4090 (GBS313) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 45 (lane 5; MW 41kDa).

GBS313-His was purified as shown in Figure 204, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1334

A DNA sequence (GBSx1418) was identified in *S.agalactiae* <SEQ ID 4093> which encodes the amino acid sequence <SEQ ID 4094>. This protein is predicted to be DNA polymerase III alpha subunit (dnaE). Analysis of this protein sequence reveals the following:

Possible site: 55
 30 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1446(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 4096.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 1335

A DNA sequence (GBSx1419) was identified in *S.agalactiae* <SEQ ID 4097> which encodes the amino acid sequence <SEQ ID 4098>. This protein is predicted to be YHCF (farR). Analysis of this protein sequence reveals the following:

Possible site: 52
 45 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3316(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

-1474-

>GP:BAB04102 GB:AP001508 transcriptional regulator (GntR family)

[Bacillus halodurans]

Identities = 51/116 (43%), Positives = 79/116 (67%)

5 Query: 5 FNEKSPIYSQIAEHIKMQIVSQEIKSGDQLPTVRELAQEAGVNPNTMQRAFTELEREQMV 64
F+ PIY Q+AE +K QIV E++ G++LP+VR++ EA VNPNT+QR + ELE +V
Sbjct: 5 FHSSEPIYLQLAERVKRQIVRGELRLGEKLPVSRDMGIEANVNPNTVQRTYRELEGLKIV 64

10 Query: 65 FSQRTSGRFVTEENLLIGKIRQQVAKAELATFVNNMKKIGYKLDDEITVALDHFIKE 120
S+R G FVTED ++ IR+Q+ + E++ FV M+++GY +EI L+ ++ E
Sbjct: 65 ESKRQGQTFVTEDEQVLQAIREQMKETISHFVQGMREMGYSNEIQAGLESYLTE 120

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4099> which encodes the amino acid sequence <SEQ ID 4100>. Analysis of this protein sequence reveals the following:

15 Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2075(Affirmative) < succ>
20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 80/120 (66%), Positives = 100/120 (82%)

25 Query: 1 MAWEFNEKSPIYSQIAEHIKMQIVSQEIKSGDQLPTVRELAQEAGVNPNTMQRAFTELER 60
M+W+F EKSPIY+QIA+H+ MQI+SQEIKSGDQLPTVRE A+ AGVNPNTMQRAFTELER
Sbjct: 1 MSWKFEKSPIYQIAQHVMQIISQEIKSGDQLPTVREYAEIAGVNPNTMQRAFTELER 60

30 Query: 61 EGMVFSQRTSGRFVTEENLLIGKIRQQVAKAELATFVNNMKKIGYKLDDEITVALDHFIKE 120
EGMV+SQRT+GRFVT+D LI + R+++A +EL +F+ NM K+G+ EI L F+KE
Sbjct: 61 EGMVVSQRTAGRFVTDQKLIARKRRELAISELSFTTMMTKMGFSHTETIIPVLTSFLKE 120

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
35 vaccines or diagnostics.

Example 1336

A DNA sequence (GBSx1420) was identified in *S.agalactiae* <SEQ ID 4101> which encodes the amino acid sequence <SEQ ID 4102>. This protein is predicted to be ABC transporter, ATP-binding protein (yhcG). Analysis of this protein sequence reveals the following:

40 Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2757(Affirmative) < succ>
45 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12735 GB:Z99108 similar to glycine betaine/L-proline
transport [Bacillus subtilis]
Identities = 87/228 (38%), Positives = 150/228 (65%), Gaps = 1/228 (0%)

50 Query: 5 LQLHHVTKKYHKHTAVNDVTVSIPGKIIGLLGPNNGSGKTTTIKMINGLLQPDKGDIVID 64
++L HV+KKY +HTAVNDV++++ +G+I GL+GPNNGSGK+T +KM+ GLL P G + +D
55 Sbjct: 3 IKLEHVSKKYGRHTAVNDVSITLSSGRIYGLIGPNNGSGKSTTLKMMAGLLFPTSGFVKVD 62

Query: 65 GYRPSVETKKIISYLPDTSYLQENMKIKDVVTLFEDFYNDFDSKVAYQLFEDLNLNPRER 124
+ + E + +YL + + +KD+V ++ + DF ++ Y+L ++ LNP ++

-1475-

Sbjct: 63 EEQVIREMVRQTAYLTLEDMFYPHFTVKDMVNFYQSQFPDFHTEQVYKLLNEMQLNPEKK 122

Query: 125 LKNLSKGNKEKVQLILVMSRKARLYILDEPIGGVDPAAARDYILKTIISNYSNDAS-VLIS 183
+K LSKGN+ +++++L ++R+A + +LDEP G+DP RD I+ +++S + V+I+

5 Sbjct: 123 IKKLSKGNRRLKIVLALARRADVILLDEPFSGLDPMVRDSIVNSLVSYIDFEQQIVVIA 182

Query: 184 THLISDIEPILDEVIFLKEGEIDLQGNADDLRREEHNCISIDALFRERFK 231
TH I +IE +LDEVI L GE Q +D+RE+ S+ F+ + +

10 Sbjct: 183 THEIDEIETLLDEVIIILANGEKVAQREVEDIREQEGMSVLQWFKSKME 230

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4103> which encodes the amino acid sequence <SEQ ID 4104>. Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1983(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 171/231 (74%), Positives = 200/231 (86%)

Query: 1 MTQLQLHHVTKKYHKHTAVNDVTVSIPITGKIIGLLGPNNGSGKTTIIMINGLLQPDKGD 60
M LLQLHHV+K Y + A++D+T++IP GKIIGLLGPNNGSGKTT+IK+INGLLQP+KG+

25 Sbjct: 1 MAHLQLHHVSKSYREKKAIDDLTITIPNGKIIGLLGPNNGSGKTTLIKLINGLLQPNKGE 60

Query: 61 IVIDGYRPSVETKKIISYLPDTSYQENMKIKDVVTLFEDFYNDPDSKVAYQLFEDLNLN 120
IVIDGYRP VETKKIISYLPDT+YL ENM+IKD++ F DFY+DFD A L DL L+

30 Sbjct: 61 IVIDGYRPCVETKKIISYLPDPTYLNENMRIKDMLEFFSDFYSDFDKSKATSLLRDLLELD 120

Query: 121 PRERLKNLSKGNKEKVQLILVMSRKARLYILDEPIGGVDPAAARDYILKTIISNYSNDASV 180
P +R K LSKGNKEKVQLILVMSRKARLY+LDEPIGGVDPAAARDYILKTI++Y +ASV

35 Sbjct: 121 PEDRFKTLKSGNKEKVQLILVMSRKARLYVLDEPIGGVDPAAARDYILKTIINSYCENASV 180

Query: 181 LISTHLISDIEPILDEVIFLKEGEIDLQGNADDLRREEHNCISIDALFRERFK 231
+ISTHLISDIEPILDEVIFLK+G + L GNADDLR+E+ SID+LFRE +K

Sbjct: 181 IISTHLISDIEPILDEVIFLKQGRFLSGNADDLRQEYQQSIDSLPRETYK 231

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1337

A DNA sequence (GBSx1421) was identified in *S.agalactiae* <SEQ ID 4105> which encodes the amino acid sequence <SEQ ID 4106>. Analysis of this protein sequence reveals the following:

Possible site: 48
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -15.39	Transmembrane	120 - 136 (103 - 146)
INTEGRAL	Likelihood = -9.98	Transmembrane	55 - 71 (47 - 79)
INTEGRAL	Likelihood = -9.45	Transmembrane	22 - 38 (15 - 43)
INTEGRAL	Likelihood = -6.05	Transmembrane	192 - 208 (187 - 218)
INTEGRAL	Likelihood = -4.94	Transmembrane	230 - 246 (228 - 253)
INTEGRAL	Likelihood = -4.78	Transmembrane	157 - 173 (155 - 175)
INTEGRAL	Likelihood = -1.44	Transmembrane	103 - 119 (103 - 119)

----- Final Results -----
bacterial membrane --- Certainty=0.7156(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

-1476-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4107> which encodes the amino acid sequence <SEQ ID 4108>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
5   INTEGRAL    Likelihood =-11.52    Transmembrane  190 - 206 ( 187 - 215)
    INTEGRAL    Likelihood =-10.67    Transmembrane  121 - 137 ( 104 - 141)
    INTEGRAL    Likelihood = -5.73    Transmembrane   63 -  79 (  59 -  82)
    INTEGRAL    Likelihood = -4.83    Transmembrane  158 - 174 ( 156 - 181)
10  INTEGRAL    Likelihood = -1.38    Transmembrane  232 - 248 ( 232 - 248)
    INTEGRAL    Likelihood = -0.85    Transmembrane  104 - 120 ( 104 - 120)

----- Final Results -----
    bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
    bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
15  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 116/267 (43%), Positives = 165/267 (61%), Gaps = 13/267 (4%)
20  Query: 1  MFGKLLKYELKSVGKWYLTNAAVLLVSIILGLVLKALG-----GNFSTDTNSTSAQIFT 55
    MFGKLLKYE +S+GKWY LNA V+ ++ IL +K G F TN ++
    Sbjct: 1  MFGKLLKYEFERSIGKWYFALNAFVIAIAAILSFITIKLFAQSNSDGLFGVLTN----KMLP 56

25  Query: 56  IILVLLLAMVISGSLSTLAIIIKRFYSNIFGRQGYLTLTLPVTTNQIICSKLLASLLWS 115
    + L L +I+GSLLSTL IIKRF ++FG +GYLTLTLPV ++QII SKLLAS + S
    Sbjct: 57  LTLGLTFGSLIAGSLSTLLIIKRFKSVFGWEGYLTLTLPVNSHQIILSKLLASFICS 116

30  Query: 116 IFNIFIVIIGIILVILPLVGIGQFVVAFPPEIYKIISSSNAPLFIAYFFLSYVAGTLLIYL 175
    +FN I+ I +VI+P+ I + + F +K+ N +AY LS LLIYL
    Sbjct: 117 VFNTIILAFIAIAIVIVPMFNINELLEGGFFNSFKMDYFINMLTVLAYVLLSTFTSILLIYL 176

35  Query: 176 SIAVGQLFTNKRVLMGIVSYFGISLLITFLTLIIDSIFHIDLFNSHANA-TFSQPVLly- 233
    SI++GQLF+N+R LM ++YF + +LI+ + S HI N+ A++ F++ +Y
    Sbjct: 177 SISIGQLFSNRRGLMAFIAYFILVILISVAATYVHS--HIFNINTSADSFPFTEQKTIYL 234

40  Query: 234 NILVSIVEIAIFMYMLTHSIIKYKLNQ 260
    IL +E+ +FY+ T+ IIK KLN+Q
    Sbjct: 235 LILEQFIEMIMFYIATNFIKNKLNQ 261

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1338

A DNA sequence (GBSx1422) was identified in *S.agalactiae* <SEQ ID 4109> which encodes the amino acid sequence <SEQ ID 4110>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
50  bacterial cytoplasm --- Certainty=0.5890(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside  --- Certainty=0.0000(Not Clear) < succ>

```

The protein is similar to ORF24 from *S.faecalis*.

55 No corresponding DNA sequence was identified in *S.pyogenes*.

-1477-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1339

5 A DNA sequence (GBSx1423) was identified in *S.agalactiae* <SEQ ID 4111> which encodes the amino acid sequence <SEQ ID 4112>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3316(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 The protein is similar to ORF23 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1340

20 A DNA sequence (GBSx1424) was identified in *S.agalactiae* <SEQ ID 4113> which encodes the amino acid sequence <SEQ ID 4114>. Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4256(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein is similar to ORF22 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1341

35 A DNA sequence (GBSx1425) was identified in *S.agalactiae* <SEQ ID 4115> which encodes the amino acid sequence <SEQ ID 4116>. Analysis of this protein sequence reveals the following:

Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood =-13.37 Transmembrane 62 - 78 (55 - 84)
INTEGRAL Likelihood = -8.44 Transmembrane 19 - 35 (14 - 41)

40 ----- Final Results -----
bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein is similar to ORF21 from *S.faecalis*.

-1478-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4117> which encodes the amino acid sequence <SEQ ID 4118>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2444(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 54/236 (22%), Positives = 95/236 (39%), Gaps = 12/236 (5%)

Query: 204 KDGKLRMLMKNVWEYDKLPHMLIAGGTGGGKTYFILTLIEALLHTDSKLYILDPKN--- 259

+ GK+ ++K+ DK H IAG +G GK Y LT ++L S L I+DPK

Sbjct: 14 QQGKIPVIKHFEINLDKGSWAIAGNSGSKPY-ALTYFLSVLPKPSGLIIDI PKFDTPS 72

Query: 260 --ADLADLGSVMANVYRKEDLLSCIETFYEEEMKRSEEMQMKNYKTGKNYAYLGLEAH 317

A + + + K D +S + + ++ + + + +L +

Sbjct: 73 QWARENKIAVIHPVENHSKSDVFSQVNEQLNQCATLIQKQAILYDNPNHQFTHLTI--- 129

Query: 318 FLIFDEYVAFMEMLGTKEAVMNKLKQIVMLGRQAGFFLILACORPDAKYLGDGIRDQF 377

+ DE +A E + A + L QI +LG L L Q R D + +R+Q

Sbjct: 130 --VIDEVLALSEGVMKNKEAFFSLLSQIALLGHATKIHLFLGSQRFDHNTIPISVREQL 187

Query: 378 NFRVALGRMSEMGGYMMFGSDVQKDFFLKRIKGRGYVDVGTSVISEFYTPLVPKGY 433

N + +G +++ +F + + G G + V + S PL+ Y

Sbjct: 188 NVLLQIGNINQKTQFLFPDLDPGIVIPIGHGTGIIQVVDNEHSYQVLPLLCPTY 243

SEQ ID 4116 (GBS109d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 8 & 9; MW 71kDa) and in Figure 184 (lane 2; MW 71kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 11; MW 46kDa), Figure 128 (lane 4; MW 46kDa) and Figure 179 (lane 7; MW 46kDa). GBS109d-His was purified as shown in Figure 232 (lanes 7 & 8). GBS109d-GST was purified as shown in Figure 236, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1342

A DNA sequence (GBSx1426) was identified in *S.agalactiae* <SEQ ID 4119> which encodes the amino acid sequence <SEQ ID 4120>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1479-

Example 1343

A DNA sequence (GBSx1427) was identified in *S.agalactiae* <SEQ ID 4121> which encodes the amino acid sequence <SEQ ID 4122>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4469(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9793> which encodes amino acid sequence <SEQ ID 9794> was also identified.

The protein is similar to ORF20 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1344

A DNA sequence (GBSx1428) was identified in *S.agalactiae* <SEQ ID 4123> which encodes the amino acid sequence <SEQ ID 4124>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1367(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1345

A DNA sequence (GBSx1429) was identified in *S.agalactiae* <SEQ ID 4125> which encodes the amino acid sequence <SEQ ID 4126>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood =-10.77    Transmembrane    39 - 55 ( 34 - 64)
      INTEGRAL    Likelihood = -6.32    Transmembrane    16 - 32 ( 10 - 35)

----- Final Results -----
      bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein is similar to ORF19 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

-1480-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1346

5 A DNA sequence (GBSx1430) was identified in *S.agalactiae* <SEQ ID 4127> which encodes the amino acid sequence <SEQ ID 4128>. This protein is predicted to be antirestriction protein. Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2918(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 The protein is similar to ORF18 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1347

20 A DNA sequence (GBSx1431) was identified in *S.agalactiae* <SEQ ID 4129> which encodes the amino acid sequence <SEQ ID 4130>. Analysis of this protein sequence reveals the following:

Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -3.61 Transmembrane 75 - 91 (72 - 94)

25 ----- Final Results -----
bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30 The protein is similar to ORF17 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8793> and protein <SEQ ID 8794> were also identified. Analysis of this protein sequence reveals the following:

35 Lipop Possible site: -1 Crend: 4
McG: Discrim Score: -7.12
GvH: Signal Score (-7.5): -2.52
Possible site: 43
>>> Seems to have no N-terminal signal sequence
40 ALOM program count: 1 value: -3.61 threshold: 0.0
INTEGRAL Likelihood = -3.61 Transmembrane 37 - 53 (34 - 56)
PERIPHERAL Likelihood = 3.66 58
modified ALOM score: 1.22

45 *** Reasoning Step: 3

----- Final Results -----
bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-1481-

100.0/100.0% over 167aa

Enterococcus faecalis

EGAD|14977| hypothetical protein Insert characterized
 GP|532550|gb|AAB60016.1||U09422 ORF17 Insert characterized

ORF00720(187 - 690 of 990)

EGAD|14977|15011(1 - 168 of 168) hypothetical protein {Enterococcus faecalis}

GP|532550|gb|AAB60016.1||U09422 ORF17 {Enterococcus faecalis}

%Match = 50.3

%Identity = 100.0 %Similarity = 100.0

Matches = 168 Mismatches = 0 Conservative Sub.s = 0

```

120      150      180      210      240      270      300      330
L*AKYQLVFKTILIIKPMVGI*TFQERLSQPIMGFLKSSIKSVGTLILLADFLFYGVAQSATPIFYERIDYMKKIRSYTSI
|||||
MGFLKSSIKSVGTLILLADFLFYGVAQSATPIFYERIDYMKKIRSYTSI
                        10      20      30      40

360      390      420      450      480      510      540      570
WSVEKVLVLSINDFRLPFPITFTQMTWVVSLSFAVMILGNLPLSLMIEGAFLKYFGIPVAFTWFMSTKTFDGKKPYGFLKS
|||||
WSVEKVLVLSINDFRLPFPITFTQMTWVVSLSFAVMILGNLPLSLMIEGAFLKYFGIPVAFTWFMSTKTFDGKKPYGFLKS
                        60      70      80      90      100      110      120

600      630      660      690      720      750      780      810
VIAAYALRPKLT YAGKKVTLGRNQFQEAITAVRSEFYGISN*IH*KQSRLE*RRGMLCLL*ACSLQLLISKSRTESENTSA*F
|||||
VIAAYALRPKLT YAGKKVTLGRNQFQEAITAVRSEFYGISN
                        140      150      160

```

SEQ ID 8794 (GBS223) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 7; MW 18kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1348

A DNA sequence (GBSx1432) was identified in *S.galactiae* <SEQ ID 4131> which encodes the amino acid sequence <SEQ ID 4132>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4292(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9791> which encodes amino acid sequence <SEQ ID 9792> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1349

A DNA sequence (GBSx1433) was identified in *S.galactiae* <SEQ ID 4133> which encodes the amino acid sequence <SEQ ID 4134>. Analysis of this protein sequence reveals the following:

-1482-

Possible site: 16

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.21	Transmembrane	350 - 366 (345 - 368)
INTEGRAL	Likelihood = -0.32	Transmembrane	171 - 187 (171 - 188)

----- Final Results -----

bacterial membrane	---	Certainty=0.3484(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1350

15 A DNA sequence (GBSx1434) was identified in *S.agalactiae* <SEQ ID 4135> which encodes the amino acid sequence <SEQ ID 4136>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.30	Transmembrane	154 - 170 (148 - 177)
INTEGRAL	Likelihood = -10.30	Transmembrane	21 - 37 (17 - 50)
INTEGRAL	Likelihood = -10.03	Transmembrane	320 - 336 (316 - 367)
INTEGRAL	Likelihood = -7.43	Transmembrane	346 - 362 (337 - 367)
INTEGRAL	Likelihood = -7.01	Transmembrane	186 - 202 (180 - 206)
INTEGRAL	Likelihood = -5.36	Transmembrane	411 - 427 (404 - 430)
INTEGRAL	Likelihood = -1.17	Transmembrane	386 - 402 (386 - 402)

----- Final Results -----

bacterial membrane	---	Certainty=0.5118(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1351

35 A DNA sequence (GBSx1436) was identified in *S.agalactiae* <SEQ ID 4137> which encodes the amino acid sequence <SEQ ID 4138>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.6306(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1483-

Example 1352

A DNA sequence (GBSx1437) was identified in *S.agalactiae* <SEQ ID 4139> which encodes the amino acid sequence <SEQ ID 4140>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 22
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.2973(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1353

A DNA sequence (GBSx1438) was identified in *S.agalactiae* <SEQ ID 4141> which encodes the amino acid sequence <SEQ ID 4142>. Analysis of this protein sequence reveals the following:

```

20  Possible site: 42
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.3382(Affirmative) < succ>
25  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

There is also homology to SEQ ID 4144.

A related GBS gene <SEQ ID 8795> and protein <SEQ ID 8796> were also identified. Analysis of this protein sequence reveals the following:

```

30  Lipop: Possible site: -1   Crend: 3
   McG: Discrim Score:      11.12
   GvH: Signal Score (-7.5): 0.27
   Possible site: 24
35  >>> Seems to have a cleavable N-term signal seq.
   ALOM program   count: 0 value:  4.19 threshold:  0.0
   PERIPHERAL Likelihood =  4.19      69
   modified ALOM score:  -1.34

40  *** Reasoning Step: 3

   ----- Final Results -----
   bacterial outside --- Certainty=0.3000(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

100.0/100.0% over 332aa
                                     Enterococcus faecalis
   EGAD|36209| hypothetical protein Insert characterized
50  GP|532547|gb|AAB60019.1||U09422 ORF14 Insert characterized

   ORF00727(301 - 1299 of 1599)
   EGAD|36209|37602(1 - 333 of 333) hypothetical protein {Enterococcus
   faecalis}GP|532547|gb|AAB60019.1||U09422 ORF14 {Enterococcus faecalis}

```

-1484-

%Match = 61.7
 %Identity = 100.0 %Similarity = 100.0
 Matches = 333 Mismatches = 0 Conservative Sub.s = 0

```

5      249      279      309      339      369      399      429      459
      CSKSTTTTKYKK*TTNQNRHH*ESR*ETMKLKTIVIGGSGFLMVFSLLLFVAILFSDEQDSGISNIHYGGVNVSAEVLAH
      |||
      MKLKTIVIGGSGFLMVFSLLLFVAILFSDEQDSGISNIHYGGVNVSAEVLAH
      10      20      30      40      50

10     489      519      549      579      609      639      669      699
      KPMVEKYAKEYGVVEEYVNILLAI IQVESGGTAEDVMQSSSLGLPPNSLSTEESIKQGVKYFSELLASSERLSVDLESVI
      |||
      KPMVEKYAKEYGVVEEYVNILLAI IQVESGGTAEDVMQSSSLGLPPNSLSTEESIKQGVKYFSELLASSERLSVDLESVI
      70      80      90      100      110      120      130

15     729      759      789      819      849      879      909      939
      QSYNYGGGFLGYVANRGNKYTFBLAQSFKEYSGGEKVSYPNPIAIPINGGWRYNYGNMFYVQLVTQYLVTTTFDDDTVQ
      |||
      QSYNYGGGFLGYVANRGNKYTFBLAQSFKEYSGGEKVSYPNPIAIPINGGWRYNYGNMFYVQLVTQYLVTTTFDDDTVQ
      150      160      170      180      190      200      210

20     969      999      1029      1059      1089      1119      1149      1179
      AIMDEALKYEGWRYVYVGASPTTSFDCSGLTQWTYKAGINLPRTAQQQYDVTQHIPLSEAQAGDLVFFHSTYNAGSYIT
      |||
      AIMDEALKYEGWRYVYVGASPTTSFDCSGLTQWTYKAGINLPRTAQQQYDVTQHIPLSEAQAGDLVFFHSTYNAGSYIT
      230      240      250      260      270      280      290

25     1209      1239      1269      1299      1329      1359      1389      1419
      HVGIYLGNNRMFPHAGDPIGYADLTSPYWQQHLVGAGRIKQ*ERKI***NLEKIRIKKNRYQRKRNLSIRSILIKRL*LP
      |||
      HVGIYLGNNRMFPHAGDPIGYADLTSPYWQQHLVGAGRIKQ
      310      320      330
  
```

35 SEQ ID 8796 (GBS155) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 24 (lane 10; MW 38kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 7; MW 62kDa).

The GBS155-GST fusion product was purified (Figure 111; see also Figure 198, lane 74) and used to immunise mice (lane 1 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1354

45 A DNA sequence (GBSx1439) was identified in *S.agalactiae* <SEQ ID 4145> which encodes the amino acid sequence <SEQ ID 4146>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -8.60    Transmembrane    37 - 53 ( 35 - 55)

50  ----- Final Results -----
        bacterial membrane --- Certainty=0.4439(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
  
```

55

-1485-

A related GBS nucleic acid sequence <SEQ ID 9789> which encodes amino acid sequence <SEQ ID 9790> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1355

A DNA sequence (GBSx1440) was identified in *S.agalactiae* <SEQ ID 4147> which encodes the amino acid sequence <SEQ ID 4148>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -0.00    Transmembrane 391 - 407 ( 391 - 407)

----- Final Results -----
      bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9787> which encodes amino acid sequence <SEQ ID 9788> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4149> which encodes the amino acid sequence <SEQ ID 4150>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2027(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 183/669 (27%), Positives = 305/669 (45%), Gaps = 63/669 (9%)

Query: 7   KIINIGVLAHVDAKKTTLTESLLYNSGAITELGSVDKGTTRTDNTLLERQRGITITQTGIT 66
          K NIG++AHVDACKTT TE +LY +G I ++G +G ++ D E++RGITI + T
Sbjct: 9   KTRNIGIMAHVDAGKTTTTERILYYTGKIHKIGETHEGASQMDWMEQEQERGITTITSAAT 68

Query: 67  SFQWENTKVNIIIDTPGHMDFLAEVYRSLSVLDGAILLISAKDGVQAQTRILFHALRKMGI 126
          + QW+ +VNIIDTPGH+DF EV RSL VLDGA+ ++ ++ GV+ QT ++ + G+
Sbjct: 69  TAQWDGHRVNIIDTPGHVDFTEVQVRSRLRVLDGAVTVLDSQSGVEPQTETVWRQATEYGV 128

Query: 127 PTIFFINKIDQNGIDLSTVYQDIKEKLSAEI-----VIKQKVELYPN 168
          P I F NK+D+ G D Q + ++L A +IK K E+Y N
Sbjct: 129 PRIVFANKMDKIGADFLYSVQTLHDLRLQANAHPIQLPIGAEDDFRGIIDLKMKAEIYTN 188

Query: 169 MCVTNFTES---EQW-----DTVIEGNDLLLEKYMMSGKSLEALELEQESIRF 213
          T+ E E++ + V E ++DL+ KY+ G+ + EL
Sbjct: 189 DLGTDILEEDIPEEYLEQAQYREKLIIEAVAETDEDLMMKYLEGEBITNDELIAGIRKAT 248

Query: 214 HNCSLFPVYHGSAKNNIGIDNLEIVI-----TNKFYSSTHRGPSE----L 254
          N FPV GSA N G+ +++ + N + P+
Sbjct: 249 INVEFFPVLCGSAPFNKGVQLMLDAVIAYLPSPDLIPAIGVNPDTDAEEERPASDEEFP 308

Query: 255 CGNVFKIEYTKRQRLAYIRLYSGVLHLRDSVRVSEKEKI---KVTEMYTSINGELCKI 310
          FKI RL + R+YSGVL+ V + K K ++ +M+ + E I
Sbjct: 309 AALAFKIMTDPFVGRLTFFRVYSGVLNSGSYVMNTSKGKRERIGRILQMHANSRQE---I 365

```

-1486-

5 Query: 311 DRAYSGETVILQN-EFLKLNVLGDTKLLPQRKKIENPHPLLQTTVEPSKPEQREMLLDA 369
 + Y+G+I + L D K + IE P P++Q VEP ++ + A
 Sbjct: 366 ETVYAGDIAAAVGLKDTTGTGDSLTDEKAKVILESIEVPEPVIQLMVEPKSKADQDKMGVA 425

10 Query: 370 LLEISDSDFLLRYVVDSTTHEIILSFLGKVQMEVISALLQEKYHVEIELKEPTVIYME-- 427
 L ++++ DP R + T E +++ +G++ ++V+ ++ ++ VE + P V Y E
 Sbjct: 426 LQKLAEDPTFRVETNVETGETVIAGMGELHLDVLVDRMKREFKVEANVGAPQVSURETF 485

15 Query: 488 YGCEQG-LYGNVTDCKICFKYGLYSPVSTPADFRMLAPIVLEQVLKKAGTELLEPYLS 546
 G L G+ + D K G Y+ S+ F++ A + L++ K A +LEP +
 Sbjct: 546 ESMANGVLAGYPMVDVAKLYDGSYHDVDSSETAFKIAASLALKEAKSAQPAILEPMML 605

20 Query: 547 FKIYAPQEYLSRAYNDAPKYCANIVDTQLKNNVILSGEIPARCIQEYRSDLTFFFTNGRS 606
 I AP++ L + + N I+ +P + Y + L T GR
 Sbjct: 606 VTITAPEDNLGDMGHVTARRGRVDGMEAHGNSQIVRAYVPLAEMFGYATVLRSATQGRG 665

25 Query: 607 VCLTELKGY 615
 + Y
 Sbjct: 666 TFMVFDHY 674

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1356

30 A DNA sequence (GBSx1441) was identified in *S.agalactiae* <SEQ ID 4151> which encodes the amino acid sequence <SEQ ID 4152>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2530(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1357

A DNA sequence (GBSx1442) was identified in *S.agalactiae* <SEQ ID 4153> which encodes the amino acid sequence <SEQ ID 4154>. Analysis of this protein sequence reveals the following:

45 Possible site: 18
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1487-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1358

5 A DNA sequence (GBSx1443) was identified in *S.agalactiae* <SEQ ID 4155> which encodes the amino acid sequence <SEQ ID 4156>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1630(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1359

20 A DNA sequence (GBSx1444) was identified in *S.agalactiae* <SEQ ID 4157> which encodes the amino acid sequence <SEQ ID 4158>. This protein is predicted to be excisionase-related protein. Analysis of this protein sequence reveals the following:

Possible site: 40
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4481(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein is similar to transposon Tn916 from *S.faecalis*. No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1360

35 A DNA sequence (GBSx1445) was identified in *S.agalactiae* <SEQ ID 4159> which encodes the amino acid sequence <SEQ ID 4160>. This protein is predicted to be transposase. Analysis of this protein sequence reveals the following:

Possible site: 46
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4626(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein is similar the Tn1545 integrase from *S.pneumoniae* and to SEQ ID 578.

-1488-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1361

A DNA sequence (GBSx1446) was identified in *S.agalactiae* <SEQ ID 4161> which encodes the amino acid sequence <SEQ ID 4162>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood = -10.72    Transmembrane    18 - 34 ( 13 - 41)
INTEGRAL    Likelihood = -6.10     Transmembrane    58 - 74 ( 55 - 79)
10 INTEGRAL    Likelihood = -5.04     Transmembrane    97 - 113 ( 90 - 116)
INTEGRAL    Likelihood = -1.81     Transmembrane    78 - 94 ( 78 - 94)
INTEGRAL    Likelihood = -0.85     Transmembrane    145 - 161 ( 145 - 161)

----- Final Results -----
15 bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

20 >GP:AAC74820 GB:AE000270 orf, hypothetical protein [Escherichia coli K12]
    Identities = 43/174 (24%), Positives = 84/174 (47%), Gaps = 9/174 (5%)

Query: 24 LIATLVLVVLYLKL-----GILNDSNELKDLVHKYEFWGPIMIFIVAQIVQIVFPVIPGG 77
      L A L+ + +Y + +L D L+ L+ + F+G ++I+ I+ + ++PG
25 Sbjct: 24 LFACILFALVIYAIHAFGLFDLLTDLPHLQTLIRQSGFFGYSLYILLFTIATLL-LLPGS 82

Query: 78 VTTVAGFLIFGPTLGFYIYIGIIGSVILFVLVKFYGRKFVLLFM-DQKTFDKYESKLE 136
      + +AG ++FGP LG + + I + S F L ++ GR +L ++ TF E +
30 Sbjct: 83 ILVIAGGVVFGPLLGLTLLSLIAATLASSCSFLLARWLGRDLLLKYVGHSENTFQAIEKGIA 142

Query: 137 TSGYEKFFIFCMASPISPADIMVMITGLSNMSIKRFVTIIMITKPIISIIGYSYL 190
      +G + F I P+ P +I GL+ ++ + I +T I+ Y+ +
35 Sbjct: 143 RNGID-FLILTRLIPLFPYNIQNYAYGLTTIAFWPYTLISALTTLPGIVIYTFVM 195

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4163> which encodes the amino acid sequence <SEQ ID 4164>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have an uncleavable N-term signal seq
40 INTEGRAL    Likelihood = -4.30    Transmembrane    8 - 24 ( 6 - 29)
INTEGRAL    Likelihood = -0.80     Transmembrane    57 - 73 ( 57 - 73)
INTEGRAL    Likelihood = -0.00     Transmembrane    86 - 102 ( 86 - 102)

----- Final Results -----
45 bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```

50 Identities = 85/114 (74%), Positives = 101/114 (88%)

Query: 89 PTLGFYIYNYIGIIGSVILFVLVKFYGRKFVLLFMDQKTFDKYESKLETSGYEKFFIFCM 148
      P GFYIYNY+GIIIGS+ LF LVK YGRKF+LLF++ KTF KYE +LET GYEK FIFCM
55 Sbjct: 3 PVTGFYIYNYVGIIIGSIALFLLVKTYGRKFILLFVNDKTFYKYERLETGGEKLFIFCM 62

Query: 149 ASPISPADIMVMITGLSNMSIKRFVTIIMITKPIISIIGYSYLVWYGGDILKNFL 202
      ASP+SPADIMVMITGL++MS+KRFVTI++ITKPIISIIGYSYL+I+G D++ FL
Sbjct: 63 ASPVSPADIMVMITGLTMSLKRFTILLITKPIISIIGYSYLFIFGKDVISWFL 116

```

-1489-

There is also homology to SEQ ID 1728.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 1362

A DNA sequence (GBSx1447) was identified in *S.agalactiae* <SEQ ID 4165> which encodes the amino acid sequence <SEQ ID 4166>. This protein is predicted to be chlorAMPhenicol acetyltransferase (cat). Analysis of this protein sequence reveals the following:

```

10 Possible site: 28
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.4725 (Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
15                bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

20 >GP:AAA86871 GB:U19459 VAT B [Staphylococcus aureus]
    Identities = 57/130 (43%), Positives = 81/130 (61%), Gaps = 4/130 (3%)

    Query: 57 IGAFCSIAQNVT--ITGLNHPTDHTTNPFIIYKSRGFINEADRADLIDEKKNKGKVIIGND 114
              IG FC+IA+ + + G NH + ITT PF G+ + L D G ++GND
    Sbjct: 65 IGKFCIAIEGIEFIMNGANHRMNSITTYPF-NIMGNGW-EKATPSLEDLPFKGDTVVGND 122

25 Query: 115 VWIGTNVTILPSVTIGNGAILGAGSVITKDIPDYAVVAGTPAKIIKYRFSEEEITLLNAS 174
              VWIG NVT++P + IG+GAI+ A SV+TKD+P Y ++ G P++IIK RF +E I L
    Sbjct: 123 VWIGQNVTVMPGIQIGDGAIVAANSVVTKDVPVPPYRIIGGNPSRIIKRFEDELIDYLLQI 182

30 Query: 175 QWWNWSDEAI 184
              +WW+WS + I
    Sbjct: 183 KWDWSAQKI 192

```

There is also homology to SEQ ID 1944.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1363

A DNA sequence (GBSx1448) was identified in *S.agalactiae* <SEQ ID 4167> which encodes the amino acid sequence <SEQ ID 4168>. Analysis of this protein sequence reveals the following:

```

40 Possible site: 39
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.2398 (Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45                bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1490-

Example 1364

A DNA sequence (GBSx1449) was identified in *S.agalactiae* <SEQ ID 4169> which encodes the amino acid sequence <SEQ ID 4170>. This protein is predicted to be cation-transporting P-ATPase PacL. Analysis of this protein sequence reveals the following:

5 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.18	Transmembrane	873 - 889 (866 - 894)
INTEGRAL	Likelihood = -8.39	Transmembrane	257 - 273 (251 - 276)
INTEGRAL	Likelihood = -5.95	Transmembrane	67 - 83 (65 - 88)
10 INTEGRAL	Likelihood = -5.41	Transmembrane	282 - 298 (281 - 301)
INTEGRAL	Likelihood = -1.65	Transmembrane	90 - 106 (89 - 107)
INTEGRAL	Likelihood = -0.48	Transmembrane	737 - 753 (736 - 753)
INTEGRAL	Likelihood = -0.00	Transmembrane	898 - 914 (898 - 914)

15 ----- Final Results -----
 bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 10963> which encodes amino acid sequence <SEQ ID 10964> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB85991 GB:AE000912 cation-transporting P-ATPase PacL
 [Methanothermobacter thermoautotrophicus]

25 Identities = 409/922 (44%), Positives = 609/922 (65%), Gaps = 22/922 (2%)

Query: 10 TNTRFAKEELEEVFEELGTTQGGLSDEEVAVRQKKYGLNLLSEVKQESIILLFLKNFTSL 69
 T T + E+EEV + L T++ GL +E R K +G N L EVK+ +ILLFL N ++

30 Sbjct: 4 TMTAIYELEVEEVLQRLETSESGLDPOEAEKRLKIHGPNKLEEVKRRPLILFLSNLYNV 63

Query: 70 MAILLVWGGFVAIVSNSLELGLAIWMVNVINGIFSFIOEYRASQATOALEKMLPSYSRVL 129
 +A+LLW+ ++ ++ + +L +AI MV +IN +FSF QEY A +A +AL+ +LP +V+

35 Sbjct: 64 LALLLWIAAILSFTTGNVQLAVAIVMVIINALFSFWQEYEAKEAEALKNILPVMVKVI 123

Query: 130 RKGSEEKILSEQLVPGDIVLIEEGDRISADGRLIKTTDLQVNSALTGESNPIYKDSNVE 189
 R E I + +V GDI+++EEGD + AD R++++ +L+V+ SALTGES P+ K S+

40 Sbjct: 124 RASKEVLIPAADVVHGDIIILEEGDTPADARILESHNLRVDASALTGESKPVKRVSHVP 183

Query: 190 NDQSKTLIECDNMVFA GTTVSSGSATMVVTAIGMQTFQGIADLTQGMKSEKSPLORELD 249
 + + I+ +N++FAGT V+SG+ V A G T+F +IA LTQ ++ E SPLQR++

45 Sbjct: 184 RE-ADNYIDTENILFAGTQVTS GTTGRAAVFATGRDTEFSRIATLTQEVREEPSPLQRQIS 242

Query: 250 RLTKQISIIISITVGIIFLAATFFVKPEVSKSFIFALGMIVAFIPEGLLPVTLSLAMAV 309
 + I +++ +G+I FL + V+ P+ +FIFA+G+++VA +PEGILP+VTLSLA +

50 Sbjct: 243 LAARIIGALAVAMGVILFLVNLIVRLPLETAFIFAIGLMVANVPEGLLPVTLSLAASA 302

Query: 310 QRMAKEHALVKLSSVETLGATSVICSDKTGTLTQNMVTVNHLWQNGKSYQVTGLGYAPE 369
 ++MA+E+ALVK+LSSVETLG+T++IC+DKTGTLT+ EMTV +W K +VTG GY PE

55 Sbjct: 303 RKMARENALVKRLSSVETLGSTTIICTDKTGTLTRGEMTVRKIWIPIHKVIEVTGSGYRPE 362

Query: 370 GQILFEGDNICFGNSDRGDLEKLIRFAHLCSNAQVLPNDSTRSTYTVLGDPTACLNVL 429
 GQ LF G+ + + D +L+ L+R A C+++ ++ + ++VLGD TE L V

60 Sbjct: 363 GQFLFRGEPV--SHRDMAECLKLMRAATFCNDSALI---HEEGEWSVLGDSTEGALLVAA 417

Query: 430 EKSGINIQENRKFAFRLKELPFDSVRKRMTHSLGGDEKDKKISITKGAPKEILDLSY 489
 EK G + + K PR+ ELPFDS RK MT+IH G K+++ KGAPK+I+ LS+

65 Sbjct: 418 EKLGFDAEAEALKAMPRTIELPFDSRRKSMTSIEHSGS---KRVAYVKGAPKKIIGLSER 473

Query: 490 VLSDGKVIPLNKEERNKIQLANDTFAKDGLRVLAVSYCDIEGFSKEQWTQENLEQHMVFI 549
 + DG+V L+ +E+ +I +D A GLRVLA +Y ++ E +E+ +V +

70 Sbjct: 474 ISVDGRVRLHADEKERIIGIHDEMASKGLRVLAFAFRELPE-DLEVRDPGEVERDLVLV 532

-1491-

Query: 550 GLIAMS D P P R E G V R E A I D K C H A A S I R I I M V T G D Y G L T A L S I A K N I G I I R N D D A K V I S G L E 609
 G+ AM D P P R E G V+EA++ C A I R I I M+T G D Y G L T A +I A+ I G I+ + ++I G E
 Sbjct: 533 G M A A M H D P P R E G V K E A V E H C K T A G I R I I M I T G D Y G L T A E A I A R E I G I V E G - E C R I I K G K E 591

5 Query: 610 L S E M T D S Q L K K E L S G E -- V V F A R V A P E Q K Y R V V T I L Q E M G E V V A V T G D G V N D A P A L K K S D 667
 L ++ D ++ L + L + E ++ F A R P E K R + ++ L ++ E + V A + T G D G V N D A P A L + K + D
 Sbjct: 592 L D K L K D T E L R G I L A R E R N L I F A R A V P E H K M R I A S V L E D S D E I V A M T G D G V N D A P A L R K A D 651

10 Query: 668 I G V A M G V T G T D V A K E S A D M I L T D D H F A S I V H A V E E G R A V Y Q N I K K F L T Y I F N S N T P E A V P 727
 I G V A M G + G T D V A K E + A D ++ L D D + F A S I V A V E G R V Y + N I + K F + T Y I F + T E V P
 Sbjct: 652 I G V A M G - S G T D V A K E A A D I V L A D D N F A S I V T A V R E G R T V Y E N I R K F I T Y I F S H E T A E I V P 710

15 Query: 728 S A F F L F S K G F I P L P L T V M Q I L A V D L G T D M L P A L G L G V E P P E T D V M N R P P R R L T D R L L D K G 787
 F + I P L F + T + M Q I L A + D L G T D L P A L L G P E + D V M P P R ++ R L L ++
 Sbjct: 711 -- F I M M V L F S I P L P I T I M Q I L A I D L G T D T L P A L A L G R S L P E S D V M K L P P R A P S E R L I N R E 768

20 Query: 788 L L I K S F L W Y G T I E S V L A M G G F F W A H Y L R Y G N F --- T F F V A N G I P Y R E A T T M T L G A I I F S Q 844
 ++++ + L + G T I E + L M + F Y G + A + Y A T T + I + + Q
 Sbjct: 769 V I L R G Y L F T G T I E A A L I M A A Y F L V L Y -- S G G W L P G Q E L S A S D P L Y M R A T T V V F A G I V M A Q 826

Query: 845 I G M V M N S R T S Y Q S I K A L S I F G N K L I N F G I I M E I L A F L V L V Y V P L F H N L F N T A S L G L S H W L 904
 + G +++ S + T S + N + I G ++ I L +++ Y + P + F T A G + W
 Sbjct: 827 L G N L L S S Q T I R S S A L E A G L L R N R W I L A G M V F A I S V M L L V I Y L P P L Q P I F G T A P P G I L E W F 886

25 Query: 905 Y L I S C P F I M I G L D E V R K L F S S R 926
 L I I + D E + R K R
 Sbjct: 887 I L I L F T P I V F L T D E M R K F I Q R R 908

There is also homology to SEQ ID 4172.

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1365

A DNA sequence (GBSx1450) was identified in *S. agalactiae* <SEQ ID 4173> which encodes the amino acid sequence <SEQ ID 4174>. Analysis of this protein sequence reveals the following:

35 Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3740 (Affirmative) < succ>
 40 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB46979 GB:AJ243482 CSRA protein [Enterococcus faecalis]
 45 Identities = 85/132 (64%), Positives = 105/132 (79%)

Query: 2 K E T Q E E L R Q R I G H T A Y Q V T Q N S A T E H A F T G K Y D D F F E E G I Y V D I V S G E V L F S S L D K F Q S G 61
 K T + E E L + Q + Y V T Q + A T E F + G + Y D D F + + + G I Y V D I V S G E L F S S L D K + + G
 Sbjct: 3 K P T E E L K Q T L T D L Q Y A V T Q E N A T E R P F S G E Y D D F Y Q D G I Y V D I V S G E P L F S S L D K Y D A G 62

50 Query: 62 C G W P A F S K P I E N R M V T N H Q D H S H G M H R I E V R S R Q A D S H L G H V F N D G P V D A G G L R Y C I N S A 121
 C G W P + F + K P I E R V D S H G M H R + E V R S + + A D S H L G H V F D G P + G G L R Y C I N + A
 Sbjct: 63 C G W P S F T K P I E K R G V K E A D F S H G M H R V E V R S Q E A D S H L G H V F T D G P L Q E G G L R Y C I N A A 122

55 Query: 122 A L D F I P Y D Q M A K 133
 A L F + P + K
 Sbjct: 123 A L R F V F V A D L E K 134

60 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4175> which encodes the amino acid sequence <SEQ ID 4176>. Analysis of this protein sequence reveals the following:

-1492-

Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3692(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 109/142 (76%), Positives = 126/142 (87%)

Query: 3 ETQEEELRQRIGHTAYQVTQNSATEHAFTGKYDDFFEEGIYVDIVSGEVLFSLLDKFQSGC 62
 ET +EL+QRIG +Y+VTQ++ATE FTG+YD+FFE+GIYVDIVSGEVLFSLLDKF SGC
 Sbjct: 2 ETSDELKQIRIGDLSYEVTQHAATESPFTGEYDNFFEKGIIYVDIVSGEVLFSLLDKFN SGC 61

15 Query: 63 GWPAFSKPIENRMVTNHQDHSHEGMRHRIEVRSRQADSHLGHVFN DGPVDAGGLRYCINSAA 122
 GWPAFSKPIENRMVTNH D S+GM R+EV+SR+A SHLGHVF+DGP +AGGLRYCINSAA
 Sbjct: 62 GWPAFSKPIENRMVTNHDDSSYGMRRVEVKSREAGSHLGHVFS DGPKEAGGLRYCINSAA 121

20 Query: 123 LDFIPYDQMAKRGYGDYLSLFD 144
 L FIPYDQM K GY +L+LFD
 Sbjct: 122 LKFIPYDQMEKEGYAQWLT LFD 143

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 1366

A DNA sequence (GBSx1451) was identified in *Sagalactiae* <SEQ ID 4177> which encodes the amino
 acid sequence <SEQ ID 4178>. Analysis of this protein sequence reveals the following:

Possible site: 25

30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.1674(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05127 GB:AP001511 unknown [Bacillus halodurans]

40 Identities = 48/152 (31%), Positives = 77/152 (50%), Gaps = 1/152 (0%)

Query: 1 MIRRAKEKDLDPDIAELLKQILMLHHEVRPDIFHTRGSKFSKEQLKEMLIDESKPIFVYES 60
 +IR A +D ++A L Q+ H + R DIF + + + E + V+
 Sbjct: 2 IIREATVQDYEEVARLHTQVHEAHVKERGDIFRSNEPTLNPSFFQAAVQGEKSTVLV FVD 61

45 Query: 61 DEGKVVVAHLFLQLQEKRDLPK-KSFKTLTYIDDL CIDEEVRGQIGQKLMDFARQYAKKHG 119
 + K+ A+ + L + LP + KT+YI DLC+DE RG IG+ + + Y K H
 Sbjct: 62 EREKIGAYSVIHLVQTPLLP TMOQRKTVYISDL CVDETRRG GIGRLIFEAIISYGAHQ 121

Query: 120 CYNITLNVWNDNQRAVSFYBKLGFKPQQTQME 151
 I L+V++ N RA +FY LG + Q+ ME
 50 Sbjct: 122 VDAIRLDVYDFNDRAKAFYHSLGMRCQKQ TME 153

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 55 vaccines or diagnostics.

-1493-

Example 1367

A DNA sequence (GBSx1452) was identified in *S.agalactiae* <SEQ ID 4179> which encodes the amino acid sequence <SEQ ID 4180>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 52
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.3285(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9785> which encodes amino acid sequence <SEQ ID 9786> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:BAB06554 GB:AP001516 unknown conserved protein [Bacillus halodurans]
   Identities = 108/211 (51%), Positives = 149/211 (70%)

   Query: 7  EDVILNATENMVHHLKNDPSGHDWFHIVVRNLAVELAHKEGANTFICQMAALLHDIID 66
   E IL + E V +L ++ SGHDW+HI RV +A + +E + F+ Q+AAL HD+ID
20  Sbjct: 3  EQAILQSAEAWVKQLMDEYSGHDWYHIRRVTLMAKAIGEQEKVDFVVFVQIAALFHDLDID 62

   Query: 67  DKICQDSKQASYELTQWLYSQDLAIEVEHILDILENISFKAGTGLTMKTLEGQIVQDAD 126
   DK+ D + A +L W+ + + +++H +DI+ ISFK G G ++ T E +VQDAD
25  Sbjct: 63  DKLVDDPETAKQQLIDWMBAGVPSQKIDHTMDIINTISFKGGHGQSLATREAMVVQDAD 122

   Query: 127 RLDAMGAIGIARTMAYSGSKGRLIHDPNLKPRENLTL EEYRNGQDTAIHFYEKLLKLD 186
   RLDA+GAIGIART AYSG+KG+ I+DP L RE +T+EEYR+G+ TAI HFYEKL KLKD
30  Sbjct: 123 RLDALGAIGIARTFAYSGNKGQPIYDPELPPIRETMTVEEYRHGKSTAINHFYEKLFKLD 182

   Query: 187 LMNTKQGKMLAQKRHDFLELYLAEFYAEWNG 217
   LMNT+ GK LA++RH F+E ++ F +EWNG
   Sbjct: 183 LMNTETGKQLAKERHVFMEQFIERFLSEWNG 213

```

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1368

A DNA sequence (GBSx1453) was identified in *S.agalactiae* <SEQ ID 4181> which encodes the amino acid sequence <SEQ ID 4182>. Analysis of this protein sequence reveals the following:

```

40  Possible site: 21
   >>> May be a lipoprotein

   ----- Final Results -----
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50  !GB:U25448 internalin [Listeria monocytogenes]
   !GB:U25448 internalin [Listeria monocytogenes]
   !GB:U25448 internalin [Listeria monocytogenes]
   !GB:U25448 internalin [Listeria monocytogenes]

55  >GP:AAA69530 GB:U25448 internalin [Listeria monocytogenes]
   Identities = 78/253 (30%), Positives = 132/253 (51%), Gaps = 2/253 (0%)

```

-1494-

- 5 Query: 531 LKQLWMTNTGITDYSFLDKMPLLEGLDISQNGIKDLSFLTQYKQSLIAAANNGITSLKP 590
L Q+ +N +TD + L + L + ++ N I D++ L L+ + NN IT + P
Sbjct: 26 LTQINFNNQITDITPLKDLTKLVILMNNNQIADITPLANLSNLTGLTFNNQITDIDP 85
- 10 Query: 591 LAELPNLQFLVLSHNNISDLTPLSNLTKLQELYLDHNNVKNLSALSGLKDLKVLDSLNNK 650
L L NL L LS N ISD++ LS LT LQ+L L N V +L L+ L+ LD+S+NK
Sbjct: 86 LKNLTNLNRLELSSNTISDISALSGLTSLQQLSLG-NQVTDLKPLANLTTLERLDISSNK 144
- 15 Query: 710 EAEQNQIKSLVLKKNQKSLKFLNVTNNQTSLEGVNNYTSLETLSVSKNKLESLDIKTPN 769
+ NQI +L L L + NQ++++ + T+L L +++N+LE + +
Sbjct: 205 DLANNQISNLAFLPGLTKLTELKLGANQISNIXPLAGLTALTNLLENQLEDISPISNL 264
- 20 Query: 770 KTVTNDLDFSHNNV 782
K +T L NN+
Sbjct: 265 KNLTYLTLYFNNI 277
Identities = 91/300 (30%), Positives = 141/300 (46%), Gaps = 42/300 (14%)
- 25 Query: 519 INDMTPVLQFKKLKQLWMTNTGITDYSFLDKMPLLEGLDISQNGIKD---LSFLTQYKQL 575
I D+TP+ L L + N ITD L + L L++S N I D LS LT +QL
Sbjct: 58 IADITPLANLSNLTGLTFNNQITDIDPLKNTNLNRLELSSNTISDISALSGLTSLQQL 117
- 30 Query: 576 SLIAAANNGITSLKPLA-----ELPNLQFLVLSHNNISDLTPL 613
SL N +T LKPLA +L NL+ L+ ++N ISD+TPL
Sbjct: 118 SL----GNQVTDLKPLANLTTLERLDISSNKVSDISVLAKLTNLESLIATNNQISDITPL 173
- 35 Query: 614 SNLTKLQELYLDHNNVKNLSALSGLKDLKVLDSLNNKSADLSTLK-TTSLETLLNLTNT 672
LT L EL L+ N +K++ L+ +L LDL+NN+ ++L+ L T L L L
Sbjct: 174 GILTNDLDELNLGNQKDIGTFLASLTNLTDLANNQISNLAFLPGLTKLTELKLGANQI 233
- 40 Query: 673 SNLSFLKQKPNKVSNTINNKLASLDGIEESDEIVKVEAEQNQIKSLVLKKNQKSLKFLN 732
SN+ L ++NL +N +L + I + + N I + + L+ L
Sbjct: 234 SNIXPLAGLTALTNLLENQLEDISPISNLKNTLYLTLYFNNISDISPVSSLTQLQRLF 293
- 45 Query: 733 VTNNQTSLEGVNNYTSLETLSVSKNKLESLDIKTPNKTVTNDLDFSHNNVPTSQLKNEK 792
NN+++ + + N T++ LS N++ L TP +T + +QL LN++
Sbjct: 294 FYNKVSVDVSSLANLTNINWLSAGHNQISDL---TPLANLTRI-----TOLGLNDQ 341
Identities = 73/253 (28%), Positives = 124/253 (48%), Gaps = 4/253 (1%)
- 50 Query: 540 GITDYSFLDKMPLLEGLDISQNGIKDLSFLTQYKQSLIAAANNGITSLKPLAELPNLQF 599
GI L+ + L ++ S N + D++ L +L I NN I + PLA L NL
Sbjct: 13 GIKSIDGLEYLNNLTQINFNNQITDITPLKDLTKLVILMNNNQIADITPLANLSNLTG 72
- 55 Query: 600 LVLSHNNISDLTPLSNLTKLQELYLDHNNVKNLSALSGLKDLKVLDSLNNKSADLSTLKT 659
L L +N I+D+ PL NLT L L L N + ++SALSG L+ L L N +
Sbjct: 73 LTLFNNQITDIDPLKNTNLNRLELSSNTISDISALSGLTSLQQLSLGNQVTDLKPLANL 132
- 60 Query: 660 TSLETLLNLTNTSNTLSFLKQKPNKVSNTINNKLASLDGIEESDEIVKVEAEQNQIKSL 719
T+LE L ++ S++S L + + +L N +++ + + + ++ GNQ+K +
Sbjct: 133 TTLERLDISSNKVSDISVLAKLTNLESLIATNNQISDITPLGILTNDLDELNLGNQKDI 192
- 65 Query: 720 VLKKNQKSLKFLNVTNNQTSLEGVNNYTSLETLSVSKNKLESLDIKTPNKTVTNDLDFSH 779
+L L++ NNQ+++L + T L L + N++ ++ +TNL+ +
Sbjct: 193 GTLASLTNLTDLANNQISNLAFLPGLTKLTELKLGANQISNIXPLAGLTALTNLLENE 252
- 70 Query: 780 NNV----PTSQK 788
N + P S LK
Sbjct: 253 NQLEDISPISNLK 265
Identities = 56/209 (26%), Positives = 115/209 (54%), Gaps = 2/209 (0%)
- 75 Query: 575 LSLIAAANNGITSLKPLAELPNLQFLVLSHNNISDLTPLSNLTKLQELYLDHNNVKNLSA 634
++ + A GI S+ L L NL + S+N ++D+TPL +LTKL ++ +++N + +++
Sbjct: 4 VTTLQADRLGIKSIDGLEYLNNLTQINFNNQITDITPLKDLTKLVILMNNNQIADITP 63

-1495-

Query: 635 LSGKKDLKVLDSLNNKSADLSTLKT-TSLETLILLNETNTSNLSFLKQNPKVSNTINNNAK 693
 L+ +L L L NN+ D+ LK T+L L L+ S++S L + L++ N +
 Sbjct: 64 LANLSNLTGLTLFNNQITDIDPLKNLTNLNRLELSSNTISDISALSGLTSLQQLSLGN-Q 122

5 Query: 694 LASLDGIEESDEIVKVEAEGNQIKSLVLKKNQGSLLKFLNVTNNQLTSLEGVNNYTSLETL 753
 + L + + +++ N++ + + K +L+ L TNNQ++ + + T+L+ L
 Sbjct: 123 VTDLKPLANLTTLERLDISSNKVSDISVLAKLTNLES LIATNNQISDITPLGILTNLDEL 182

10 Query: 754 SVSKNKLESLDIKTPNKTVTNLDFSHNNV 782
 S++ N+L+ + +T+LD ++N +
 Sbjct: 183 SLNGNQLKDIGTLASLTNLTLDDLANNQI 211
 Identities = 61/228 (26%), Positives = 118/228 (51%), Gaps = 3/228 (1%)

15 Query: 483 LATVTKINIGQRTNPFQRFGLSLMPNIEVLGIGFTPINDMTPVLQFKKLKQLWMTNTGIT 542
 L ++ ++++G + + L+ + +E L I ++D++ + + L+ L TN I+
 Sbjct: 111 LTSLQQLSLGNQVTDLKP--LANLTTLERLDISSNKVSDISVLAKLTNLES LIATNNQIS 168

20 Query: 543 DYSFLDKMPLLEGLDISQNGIKDLSFLT KYKQLSLIAAANNGITSLKPLAELPNLQFLVL 602
 D + L + L+ L ++ N +KD+ L L+ + ANN I++L PL L L L L
 Sbjct: 169 DITPLGILTNLDEL SLNGNQLKDIGTLASLTNLTLDDLANNQISNLAPLPGLTKLTELKL 228

25 Query: 603 SHNNISDLTPLSNLTKLQELYLDHNNVKNLSALSGKKDLKVLDSLNNKSADLSTLKT-TS 661
 N IS++ PL+ LT L L L+ N ++++S +S K+L L L N +D+S + + T
 Sbjct: 229 GANQISNIXPLAGLTALTNLLENQLEDISPISNLKNLTLYLYFNINISDISPVSSLT K 288

30 Query: 662 LETLLNETNTSNLSFLKQNPKVSNTINNNAKLASLDGIEESDEIVKV 709
 L+ L S++S L ++ L+ + +++ L + I ++
 Sbjct: 289 LQRLFFYNNKVSVDVSSLANLTNINWLSAGHNQISDLTPLANLTRITQL 336
 Identities = 60/286 (20%), Positives = 129/286 (44%), Gaps = 24/286 (8%)

35 Query: 369 SNKLSDEDDQKKLIYLAEKLG LNPNQIEVLTSSEDSIIIFKYPHDDHSHTIASKDIEIGKPI 428
 +N+++D D K + +L L+ N I +++ G + + + +G +
 Sbjct: 77 NNQITDIDPLKNLTNLNRLELSSNTISDISALSG-----LTSLQQLSLGNQV 123

40 Query: 429 PDGHHDSHAKDKVGMATLKQIGFDEIIQDILHADAPTFPPSNETNPEKMRQW--LATV 486
 D K + TL+++ + DI T S ++ L +
 Sbjct: 124 TD-----LKPLANLTTLERLDISSNKVSDISVLAKLTNLES LIATNNQISDITPLGIL 176

45 Query: 487 TKIN-IGQRTNPFQRFGLSLMPNIEVLGIGFTPINDMTPVLQFKKLKQLWMTNTGITY 544
 T ++ + N + G L+ + N+ L + I+++ P+ KL +L + I++
 Sbjct: 177 TNLDELSLNGNQLKDIGTLASLTNLTLDDLANNQISNLAPLPGLTKLTELKLGANQISNI 236

50 Query: 545 SFLDKMPLLEGLDISQNGIKDLSFLT KYKQLSLIAAANNGITSLKPLAELPNLQFLVL 604
 L + L L++++N ++D+S ++ K L+ + N I+ + P++ L LQ L +
 Sbjct: 237 XPLAGLTALTNLLENQLEDISPISNLKNLTLYLYFNINISDISPVSSLTQLQRLFFYN 296

55 Query: 605 NNISDLTPLSNLTKLQELYLDHNNVKNLSALSGKKDLKVLDSLNNK 650
 N +SD++ L+NLT + L HN + +L+ L+ + L L++ +
 Sbjct: 297 NKVSDVSSLANLTNINWLSAGHNQISDLTPLANLTRITQLGLNDQE 342

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4183> which encodes the amino acid sequence <SEQ ID 4184>. Analysis of this protein sequence reveals the following:

Possible site: 21

55 >>> May be a lipoprotein

----- Final Results -----

60 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAA69530 GB:U25448 internalin [*Listeria monocytogenes*]
 Identities = 88/279 (31%), Positives = 149/279 (52%), Gaps = 2/279 (0%)

65

-1496-

Query: 419 LPNLETLGIGFTPIKDISPVLQFKKLKQLLMTKTGVTDYRFLDNMPQLEGIDISQNNLKD 478
 L + TL IK I + L Q+ + +TD L ++ +L I ++ N + D
 Sbjct: 1 LDXVTTLQADRLGIKSIDGLEYLNNLTQINFNNQLTDITPLKDLTKLVDILMNNNQIAD 60

5 Query: 479 ISFLSKYKNLTLVAAADNGIEDIRPLGQLPNLKFLVLSNNKISDLSPLASLHQLQELHID 538
 I+ L+ NLT + +N I DI PL L NL L LS+N ISD+S L+ I LQ+L +
 Sbjct: 61 ITPLANLSNLTGLTLFNNQITDIDPLKNTLNLRLELSSNTISDISALSGLTSLQQLSL- 119

10 Query: 539 NNQITDLSFVSHKESLTVVLSRNADVLDLTL-QAPKLETLMVNDTKVSHLDFLKNNPNL 597
 NQ+TDL P+++ +L +D+S N D++ L + LE+L+ + ++S + L NL
 Sbjct: 120 GNQVTDLKLPLANLTTLERLDISSNKVSDISVLAKLTNLES LIATNNQISDITPLGILTNL 179

15 Query: 598 SSSLINRAQLQSLEGIEASSVIVRVEAEGNQIKSLVLKDKQGSFLDVTGNQLTSLEGV 657
 LS+N QL+ + + + + + + NQI +L LT L + NQ++++ +
 Sbjct: 180 DELSLNGNQLKDIGTLASLTNLTDLDLANNQISNLAFLPGLTKLTELKLGANQISNIXPL 239

Query: 658 NNFTALDILSVSKNQLTNVNLSPKNTVTNIDISHNNIS 696
 TAL L +++NQL +++ K +T + + NNIS
 Sbjct: 240 AGLTALTNLELNENQLEDISPISNLKNTVLTLYFNIS 278

20

An alignment of the GAS and GBS proteins is shown below.

Identities = 346/753 (45%), Positives = 472/753 (61%), Gaps = 63/753 (8%)

25 Query: 187 SRLGNQSNSHYRVNSSK-----IAGLHYPTSNGFLFNGRG-IKGTTPTGILVEHHNH 237
 SR G SN + SK +AG+ +PT +GF+ I T GI+V+H H
 Sbjct: 38 SRKGMTSNKIKPIKSKTKNTHKGVAGVDFPTDDGFILTKDKSKILSKTDQGI VVDHGH 97

30 Query: 238 LHFISFADLRKGGW-----GSIADRYQPQKKADSKQSPSSKKPRTENTLPKDI--KDK 289
 HFI +ADL+ + G+ + ++A S+ S + P DI +D
 Sbjct: 98 SHFIFYADLKSPFEYLIPIKGASLAKPAVAQRAASQGTSKVADPHHHYEFNPADIVAEDA 157

35 Query: 290 LAYLARE---LHLDI-----SRIRVLKTLNGEIGFEYPHDDHT 324
 L Y R H + S + T NG G +P D
 Sbjct: 158 LGYTVRHDDHFHYILKSSLSGQTQAQAKQVATRLPQTSSLVSTATANGIPGLHFPSTSDGF 217

40 Query: 325 HVIMAKDIDLSKPIPNPHHDEDEH-----HKGHHHD---ESDHKHEEHEHTK 368
 + ++K HD H H +D +++ E H+ +
 Sbjct: 218 QFNGQIVGVTKDSILVDHGHLPISFADLRQGGWAHVADQYDPAKKAEPKPAETHQTPE 277

45 Query: 369 SNKLSDEDQKLIYLAELKGLNPNQIEVLTSDEGSIIFKYPHDDHSHTIASKDIEIGKPI 428
 ++ E Q+KL YLAELKGL++P+ I+ + ++DG + +YPH DH+H + DIEIGK I
 Sbjct: 278 LSEREKEYQEKLAYLAELKGLIDPSTIKRVETQDGKLGLEYPHHDHVLMLSDIEIGKDI 337

50 Query: 429 PDGH---HDHSHAKDKVGMATLKQIGFDDEIIQDILHA-DAPTPFPSNETNPEKMRQWLA 484
 PD H H K KVM TL+ +GFD+E+I DI+ DAPTPFPSNE +P M++WLA
 Sbjct: 338 PDPHAIEHARELEKHKVGMDFLRALGFDEEVILDIVRTHDAPTPFPSNEKDPNMKEWLA 397

55 Query: 485 TVTKINIGORTNPFRFGLSLMPNIEVLGIGFTPIINDMTPVLQFKKLKQLMWTNIGITDY 544
 TV K+++G R +P QR GLSL+PN+E LGIGFTPI D++PVLQFKKLKQL MT TG+TDY
 Sbjct: 398 TVIKLDLGSRKDPLQRKGLSLPNLETLGIGFTPIKDISPVLQFKKLKQLLMTKTGVTDY 457

60 Query: 545 SFLDKMPLLEGLDISQNGIKDLSFLTQYKQLSLIAAANNGITSLKPLAELPNLQFLVLSH 604
 FLD MP LEG+DISQN +KD+SFL+KYK L+L+AAA+NGI ++PL +LNL+FLVLS+
 Sbjct: 458 RFLDNMPQLEGIDISQNNLKDISFLSKYKNLTLVAAADNGIEDIRPLGQLPNLKFLVLSN 517

65 Query: 605 NNISDLTPLSNLTKLQELYLHNNVKNLSALSGKKDLKVLDSLNNKSADLSTLKTTSLET 664
 N ISDL+PL++L +LQEL++D+N + +LS +S K+ L V+DLS N DL+TL+ LET
 Sbjct: 518 NKISDLSPLASLHQLQELHIDNNQITDLSFVSHKESLTVVLSRNADVLDLTLQAPKLET 577

Query: 665 LILNETNTSNLSFLKQNPVSNLTINNAKLASLDGIEESDEIVKVEAEGNQIKSLVLKKN 724
 L++N+T S+L FLK NP +S+L+IN A+L SL+GIE S IV+VEAEGNQIKSLVLK+K
 Sbjct: 578 LMVNDTKVSHLDFLKNNPNLSSSLINRAQLQSLEGIEASSVIVRVEAEGNQIKSLVLKDK 637

Query: 725 QGSLKFLNVTNNQLTSLEGVNNYTSLETLSVSKNKLESLDIKTPNKTVTNLD FSHNNVPT 784
 QGSL FL+VT NQLTSLEGVNN+T+L+ LSVSKN+L ++++ PNKTVTN+D SHNN+
 Sbjct: 638 QGSLTFLDVTGNQLTSLEGVNNFTALDILSVSKNQLTNVNLSPKNTVTNIDISHNNISL 697

-1497-

Query: 785 SQLKLNENIPEAVAKNFPVVEGSMVGNGLAEKAAMASKEDKQVSD-NTNHQKNTEKS 843
 + LKLINE++IPEA+AKNFPV EGSVMVGN+ EKAAMA+K + + +H N +
 Sbjct: 698 ADLKLNEQHIPEAIAKNFPVVEGSMVGNGTAEKKAAMATKAKESAQEAESHYNNHT 757

5 Query: 844 AQANADSKKENPKTHDEHHDHEETDHAHVGH 876
 + E+ D H+HE+ + A +H
 Sbjct: 758 YEDEEGHAHEHRDKDDHDHEHEDNEAKDEQNH 790

10 SEQ ID 4182 (GBS84) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 9; MW 97.6kDa).

GBS84-His was purified as shown in Figure 194, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1369

15 A DNA sequence (GBSx1454) was identified in *S.galactiae* <SEQ ID 4185> which encodes the amino acid sequence <SEQ ID 4186>. This protein is predicted to be GTP-binding protein lepa (lepA). Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence
 20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1962(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14493 GB:Z99117 GTP-binding protein [Bacillus subtilis]
 Identities = 464/603 (76%), Positives = 540/603 (88%)

30 Query: 8 KRQEKIRNFSIIAHIDHGKSTLADRILEKTETVSSREMQAQLDSDMLERERGITIKLNA 67
 +RQ +IRNFSIIAHIDHGKSTLADRILEKT ++ REM+ QLLDSMDLERERGITIKLN+
 Sbjct: 9 ERQSRIRNFSIIAHIDHGKSTLADRILEKTSAITQREMKQQLDSDMLERERGITIKLNS 68

35 Query: 68 IELNYTAKDGETYIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAAQGIEAQTLANVYL 127
 ++L Y AKDGE YIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAAQGIEAQTLANVYL
 Sbjct: 69 VQLKYKAKDGEEYIFHLIDTPGHVDFTYEVSRSLAACEGAILVVDAAQGIEAQTLANVYL 128

40 Query: 128 ALDNDLEILPVINKIDLPAADPERVRAEVEDVIGLDASEAVLASAKAGIGIEEILEQIVE 187
 ALDNDLEILPVINKIDLPA+PERVR EVEDVIGLDASEAVLASAKAGIGIEEILEQIVE
 Sbjct: 129 ALDNDLEILPVINKIDLPSAEPERVRQEVEDVIGLDASEAVLASAKAGIGIEEILEQIVE 188

45 Query: 188 KVPAPTGEVDAPLQALIFDSVYDAYRGVILQVRIVNGMVKPGDKIQMMSNGKTFDVTEVG 247
 KVPAPTG+ +APL+ALIFDS+YDAYRGV+ +R+V G VKPG KI+MM+ GK F+VTEVG
 Sbjct: 189 KVPAPTGDPEAPLKALIFDSLNDAYRGVVAYIRVVEGTGKPGQIKMMATGKEFEVTEVG 248

50 Query: 248 IFTPKAVGRDFLATGDVGYYAASIKTVADTRVGDITITLANNPAIEPLHGYQMNMVVFAG 307
 +FTPKA + L GDVG++ ASIK V DTRVGDITIT A NPA E L GY+++NPMV+ G
 Sbjct: 249 VFTPKATPTNELTVGDVGFLTASIKNVGDTRVGDITITSAANPAEALPGYRKLNMVYCG 308

55 Query: 308 LYPIESNKYNDLREALEKLQNDASLQFEPETSQALGFGFRGFLGLHMDVIQERLERE 367
 LYPI++ KYNDLREALEKL+LND+SLQ+E ETSQALGFGFRGFLG+LHM++IQER+ERE
 Sbjct: 309 LYPIDTAKYNDLREALEKLENDSSLQYEATSQALGFGFRGFLGLHMEIIQERIERE 368

Query: 368 FNIDLINTAPSVVYHVNTTDGEMLEVSNPSEFPDPTRVDSIEEPYVKAQIMVPQEFVGAV 427
 FNIDLI TAPSV+Y V TDGE + V NPS PDP +++ +EPPYVKA +MVP ++VGAV
 Sbjct: 369 FNIDLITTAPSVIYDVYMTDGEKVVVDNPSNMPDPQKIERVEEPYVKATMVPNDYVGAV 428

Query: 428 MELAQKRKRGDFVTMDYIDNVRNVIIQIPLAEIVDFDFDKLSSTRGYASFDEISEYRR 487

-1498-

MEL Q KRG+F+ M Y+D NRV++IY +PLAEIV++FFD+LKSST+GYASFDYE+ Y+
 Sbjct: 429 MELCQGKRGNFIDMQYLDANRVSIYDMPLAEIVYEFFDQLKSSTKGVASFDYELIGYKP 488

Query: 488 SQLXKMDILLNGDKVDALSFIHKEFAYERGKLIIVDKLKKIIPRQQFEVPIQAAIGQKIV 547
 S+L KMDI+LNG+K+DALSFIVH+++AYERGK+IV+KLK++IPRQQFEV+QAAIGQKIV

Sbjct: 489 SKLVKMDIMLNGEKIDALSFIHVDYAYERGKVIVEKLKELIPRQQFEVFPVQAAIGQKIV 548

Query: 548 ARSDIKALRKNVLAKCYGGDVSRRKRLLEKQKAGKKRMKAIGSVEVPQEAFLSVLSMDDD 607
 ARS IKA+RKNVLAKCYGGD+SRKRKLEKQK GK+RMK +GSVEVPQEAFLSVLSMDDD

Sbjct: 549 ARSTIKAMRKNVLAKCYGGDISRRKRLLEKQKEGKRRMKQVGSVEVPQEAFLSVLSMDDD 608

Query: 608 DKK 610

KK

Sbjct: 609 PKK 611

A related GBS sequence was identified <SEQ ID 10775> which encodes the amino acid sequence <SEQ ID 10776>. A further related GBS nucleic acid sequence <SEQ ID 10955> which encodes amino acid sequence <SEQ ID 10956> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4187> which encodes the amino acid sequence <SEQ ID 4188>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1829(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14493 GB:Z99117 GTP-binding protein [Bacillus subtilis]
 Identities = 463/603 (76%), Positives = 542/603 (89%)

Query: 8 KRQEKIRNFSTIIAHIDHGKSTLADRILEKTETVSSREMQAQLLDSDMLERERGITIKLNA 67
 +RQ +IRNFSTIIAHIDHGKSTLADRILEKT ++ REM+ QLLDSMDLERERGITIKLN+

Sbjct: 9 ERQSRIRNFSTIIAHIDHGKSTLADRILEKTSAITQREMKQBQLLDSDMLERERGITIKLNS 68

Query: 68 IELNYTAKDGETYIFHLIDTPGHVDFTYEVSRSRLAACEGAILVVDAAQGIEAQTLANVYL 127
 ++L Y AKDGE YIFHLIDTPGHVDFTYEVSRSRLAACEGAILVVDAAQGIEAQTLANVYL

Sbjct: 69 VQLKYKAKDGEYIFHLIDTPGHVDFTYEVSRSRLAACEGAILVVDAAQGIEAQTLANVYL 128

Query: 128 ALDNDLEILPVINKIDLPAAADPERVRHEVEDVIGLDASEAVLASAKAGIGIEEILEQIVE 187
 ALDNDLEILPVINKIDLP+A+PVRV EVEDVIGLDASEAVLASAKAGIGIEEILEQIVE

Sbjct: 129 ALDNDLEILPVINKIDLPASAEPRVRQEVEDVIGLDASEAVLASAKAGIGIEEILEQIVE 188

Query: 188 KVPAPTGDVDAPLQALIFDSVYDAYRGVILQVRIVNGIVKPGDKIQQMMSNGKTFDVTVEVG 247
 KVPAPTGD +APL+ALIFDS+YDAYRGV+ +R+V G VKPG KI+MM+ GK F+VTEVG

Sbjct: 189 KVPAPTGDPEAPLKAIFDSIYDAYRGVVAYIRVVEGTVKPGQKIKMMATGKEFEVTEVG 248

Query: 248 IFTPKAVGRDFLATGDVGYYAASIKTVADTRVGDTVTLANPAKEALHGYKQMNPMVFAG 307
 +FTPKA + L GDVG++ ASIK V DTRVGDT+T A NPA+EAL GY+++NPMV+ G

Sbjct: 249 VFTPKATPTNELTVGDVGFLTASIKNVGDTRVGDTTSAANPAEEALPGYRKLNPVMYCG 308

Query: 308 IYPIESNKYNDLREALEKLQNDASLQFEPETSQALGFGFRGFLGLLHMDVIERLERE 367
 +YPI++ KYNDLREALEKL+LND+SLQ+E ETSQALGFGFRGFLG+LHM++IQER+ERE

Sbjct: 309 LYPIDTAKYNDLREALEKLENDSSLYEAETSQALGFGFRGFLGMLHMEIIQERIERE 368

Query: 368 FNIDLIMTAPSVVYHVHTTDEDMEIVSNPSEFPDPTRVAFIEEPYVKAQIMVPQEFVGAV 427
 FNIDLI TAPSV+Y V+ TD + + V NPS PDP ++ +EPPYVKA +MVP ++VGAV

Sbjct: 369 FNIDLITAPSVIYDVYMTDGEKVVDNPSNMPDPQKIERVEEPYVKATMMVPNDYVGAV 428

Query: 428 MELSQKRGRDFVTMDYIDNVRNVYIQLAEIVDFDFDKLSSTRGYASFDYDMSEYRR 487